Structural and Biochemical Studies on Cohesion Establishment

Benjamin O. Wade

University College London

and

The Francis Crick Institute

PhD Supervisor: Martin R. Singleton

A thesis submitted for the degree of

Doctor of Philosophy

University College London

September 2016

Declaration

I Benjamin Wade confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Sister-chromatid cohesion is a vital cellular process which involves the topological entrapment of sister-chromatids inside the ring of a protein complex called cohesin (Smc1, Smc3, Scc3 and Scc1). The early loss of cohesion can lead to chromosome mis-segregation, which in humans can cause the development of cancers. As a result, cells have evolved complex regulatory mechanisms to control cohesion throughout the cell cycle. Broadly, the cohesion cycle can be split into four stages loading, anti-establishment, establishment and release. The establishment of sister-chromatid cohesion occurs during S-phase with the passage of the replication fork, which is due to the acetylation of Smc3 at K112 and K113 through the action of the acetyltransferase called Eco1. Apart from this acetyltransferase, a number of other non-essential establishment factors have been identified, which include Chl1, Ctf4 and RFC^{Ctf18}. RFC^{Ctf18} is a replication factor C (RFC) like complex formed of the seven subunits Ctf18, Rfc2-5, Dcc1 and Ctf8. Apart from sister-chromatid cohesion RFC^{Ctf18} has also been shown to function in intra-S phase checkpoints.

To further improve our understanding on the establishment of sister-chromatid cohesion the structures of the heterotrimer $Ctf18^{C}$-Dcc1-Ctf8 was solved along with the acetyltransferase domain (ACT) of *X. laevis* Eco2 (xEco2) bound to two substrate peptides. The structure of $Ctf18^{C}$-Dcc1-Ctf8 revealed that Dcc1 contained three tandem winged helix (WH) domains which could bind to both ssDNA and dsDNA. These WH domains were involved in recruiting RFC^{Ctf18} to stalled replication forks. The structure of xEco2 ACT domain indicated that it had different binding mechanisms for both K112 and K113 substrates. To further investigate this, a mass spectrometry assay was employed to examine the relative rates of acetylation of K112 and K113. It was discovered that K112 acetylation occurs at a faster rate than the acetylation of K113.

Acknowledgement

Firstly, I would like to thank Dr. Martin Singleton for giving me a fantastic opportunity to work in his lab group. He has provided excellent support and encouragement throughout my PhD. His mentorship over the last four years has provided me with outstanding scientific training, and without him, I would have struggled to complete my PhD. I would also like to thank members of my thesis committee, Dr. Frank Uhlmann and Dr. Nic Tapon, who have provided constant advice and suggestions over the last four years. I would additionally like to thank Cancer Research UK, who provided the initial funding for my studentship, and the Francis Crick Institute for providing me with the funding to complete my PhD.

Many of Dr. Martin Singleton's lab group have kindly provided me technical help and advice during my thesis. Special thanks goes to Dr. William Chao who provided a helping hand with the Eco1 project. His input and enthusiasm was deeply appreciated. I would also like to thank Dr. Andrew Purkiss who provided help with data-collection at Diamond Light Source in the early hours of the morning!

I am grateful to Dr. Catarina Samora and Hon Liu as they carried out the work presented in Figure 5.6 and 5.7. As well as Dr. Celine Bouchoux who conducted the *in vivo* experiments in Figure 6.9C and D. Their contributions to this thesis provided further evidence to the *in vivo* role of both Eco1 and Dcc1. In addition, I appreciate the work conducted by the Proteomic facility as they helped with the mass spectrometry analysis presented in Figure 6.1, 6.2 and 6.3.

Finally, my friends and family also deserve my gratitude for providing support throughout my thesis. A special thanks goes to my fiancé, Josine, for listening to my complaints about poorly diffracting crystals!

Table of Contents

List of Figures

List of Tables

Abbreviations

ABC	ATP-Binding Cassette
CdLS	Cornelia de Lang Syndrome
ChIP	Chromatin Immunoprecipitation
DDC	DNA Damage Checkpoint
DRC	DNA Replication Checkpoint
DSB	Double Strand Break
dsDNA	Double Stranded DNA
EM	Electron Microscopy
FT	Fourier Transform
GNAT	GCN5 Related N-Acetyl-Transferase
HU	Hydroxyurea
LB	Lysogeny Broth
MAD	Multi-Wavelength Anomalous Dispersion
MS	Mass Spectrometry
NBD	Nucleotide Binding Domains
NCS	Non-Crystallographic Symmetry
PCR	Polymerase Chain Reaction
PDB	Protein Databank
PRM	Parallel Reaction Monitoring
RFC	Replication Factor C
RF	Restriction Free
ssDNA	Single Stranded DNA
UV	Ultra Violet
WH	Winged Helix
ZnF	Zinc Finger

Chapter 1. Introduction

Complex organisms are made up of eukaryotic cells of a diameter ranging between 10-30 μm. For growth, repair and reproduction cells are required to divide. Cell division requires genomic DNA to be copied after which, mitosis occurs. Daughter cells are made with an exact genomic copy of their parent's DNA. This is an extremely challenging process not least because higher eukaryotes can have over 3 billion base pairs of DNA, which have an estimated unfolded length of over one meter. Higher eukaryotes have made this process manageable by linking sister-chromatids together from replication until the onset of anaphase.

The protein complex called cohesin is responsible for the linkage of sister-chromatids. Cohesin was identified, initially, in yeast cells by mutants that exhibited increased rates of chromosome mis-segregation (Michaelis et al., 1997; Guacci et al., 1997). Later, these proteins were shown to be required in *Xenopus laevis* (Losada et al., 1998) and human cells (Sumara et al., 2000). The core cohesin complex is formed from two structural maintenance of chromosome (Smc) proteins, Smc1 and Smc3; a kleisin subunit, called sister-chromatid cohesion 1 (Scc1) and a protein called Scc3. This complex is totally conserved throughout evolution although higher eukaryotes have multiple versions of all four subunits. For instance, human cells have two copies of the Scc3 subunit called SA1 and SA2 (Losada et al., 2000). The nomenclature varies between organisms; for simplicity in this thesis the budding yeast naming will be used unless stated otherwise. However, a list of gene names for the cohesin core subunits and accessory subunits, in five different species, is presented below (Table 1.1).

The main role of cohesin is to provide cohesion between sister-chromatids; however, it has been discovered to have numerous other functions. For instance, during the repair of double strand breaks by homologous recombination, cohesin is required to provide a linkage between sister-chromatids (Sjögren and Nasmyth, 2001). Cohesin also has roles in transcription regulation (Horsfield et al., 2007) and the development of *X. laevis* (Dorsett and Merkenschlager, 2013).

Apart from the core complex, other cohesin 'accessory' subunits are required throughout the cell cycle for efficient sister-chromatid cohesion (discussed in detail below). These include the Scc2 and Scc4 complex, Wpl1, Pds5 and the acetyltransferase Eco1. Together these proteins work to coordinate cohesin on chromosomes in a dynamic process known as the 'cohesion cycle', which will be reviewed in detail below. Firstly, further structural information is presented on the core cohesin complex.

Cohesin core subunits	**Smc**	**Smc**	**Kleisin**	**HEAT Repeat**
Saccharomyces cerevisiae	SMC1	SMC3	SCC1	SCC3
Schizosaccharomyces pombe	pms1+	pms3+	rad21+	psc3+
Drosophila melanogaster	SMC1	Cap	Rad21	SA
Xenopus laevis	smc1	smc3	rad21	stag1
Homo Sapiens	SMC1A/B	SMC3	RAD21	STAG1

Cohesin accessory proteins	**Loading**		**Establishment**
Saccharomyces cerevisiae	SCC2	SCC3	ECO1
Schizosaccharomyces pombe	mis4+	ssl3+	eso1+
Drosophila melanogaster	Nipped-B	CG4203	san eco
Xenopus laevis	nipbl	kiaa0892	esco1
Homo Sapiens	NIPBL	KIAA0892	ESCO1/2

Cohesin accessory proteins	**Anti-establishment**	
Saccharomyces cerevisiae	PDS5	RAD61
Schizosaccharomyces pombe	pds5+	wpl1+
Drosophila melanogaster	pds5	wapl
Xenopus laevis	pds5a/b	wapal
Homo Sapiens	PDS5A/B	WAPAL

Table 1.1. Gene Names for The Cohesin Subunits.

1.1 The Core Cohesin Complex

The core cohesin complex is made up of four proteins: Smc1, Smc3, Scc1 and Scc3 (Michaelis et al., 1997). In this section, I will discuss information currently available on the structure of cohesin. I will start by reviewing Smc proteins and then describe, in more detail, the cohesin specific subunits.

1.1.1 Overall architecture of Smc proteins

All genomes from bacteria to humans have been shown to contain at least one Smc complex (Hirano, 2006). These complexes have been revealed to be fundamental to chromosome biology, and structural insights have proved key in understanding their mechanistic function. There are six Smc genes encoded in higher eukaryotes, which form three heterodimers. The three complexes are made from Smc1 and 3, which form a complex called cohesin, and is required for chromosome segregation; Smc2 and 4, which make a complex called condensin, and is involved in chromosome resolution at mitosis; and Smc5 and 6, which is currently unnamed, and functions in DNA repair pathways (Lehmann, 2005).

Individual Smc proteins can be as long as 1300 residues, and have a domain organisation which is unique to this family of proteins (Figure 1.1A). The N-terminus and C-terminus interact to form a globular domain termed the 'head'. Each head domain contains two nucleotide binding motifs, Walker A and B, which are separated by an antiparallel coiled-coil (Saitoh et al., 1994). Present between the two coiled-coils is another globular domain, which is responsible for the dimerisation of the Smc proteins, which is called the 'hinge'. A series of biochemical experiments revealed that by replacing the hinge of Smc3 with Smc1 you could form a homodimer with mutant and wild type Smc3 proteins. This suggested the hinge domain was responsible for the formation of the three different Smc complexes found in higher eukaryotes (Haering et al., 2002).The head domain of Smc proteins shows sequence similarities to a family of nucleotide binding proteins called the ATP-binding cassette (ABC) ATPases, which include Rad50; a double strand break (DSB) repair protein (Saitoh et al., 1994).

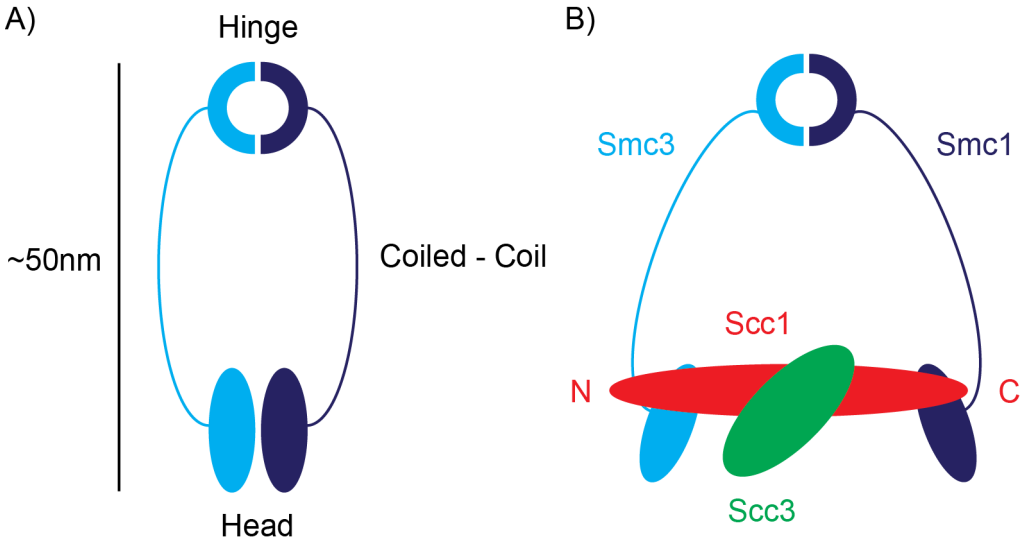


Figure 1.1. Structure of Smc Complexes.

A) The secondary structure elements of all Smc heterodimers. B) The cohesin specific subunits.

Electron microscopy (EM) analysis of MukB, a bacterial Smc protein, revealed that Smc proteins form an extended 'V shape' of ~50 nm, which can exist *in vitro* in an open or closed conformation (Melby et al., 1998). Analysis of vertebrate condensin and cohesin show different coiled-coil arrangements. Condensin coiled-coils are close together forming a 'lollypop' structure, while cohesin arms exist in an open 'ring' conformation (Anderson, 2002). To form a functional complex, the head domain of Smc complexes interact with other regulatory subunits (Figure 1). For instance, the head domain of condensin is formed from the additional kleisin subunit CAP-H, and two homologous huntingtin, elongation factor 3, A subunit and TOR (HEAT) repeat proteins called CAP-D2 and CAP-G. The Smc5 and Smc6 heterodimer binds to four subunits called Nse1, Nse2, Nse3 and Nse4 (Hirano, 2006). Recently, Nse1 and Nse3 have been shown to contain tandem winged helix (WH) domains (Palecek and Gruber, 2015).

1.1.2 The head domain of cohesin

The topological entrapment of DNA requires cohesin to form a tripartite ring (Gruber et al., 2003) made from two Smc subunits, Smc1 and Smc3, and the kleisin subunit Scc1 (Figure 1.1B). One molecule of Scc1 has been shown to bind to the Smc1 and

3 heterodimer (Haering et al., 2002). Scc1 is an elongated protein with highly conserved N- and C-terminal domains, and two conserved protease sites in the centre of the protein (see below). Scc1 completes the ring through binding to the nucleotide binding domains (NBD) of both Smc subunits; the N- and C-terminal regions of Scc1 bind to Smc3 and Smc1 respectively (Haering et al., 2002). The crystal structure of the head domain of Smc1 and Scc1 revealed that the C-terminal region of Scc1 forms a WH domain, which is able to bind to Smc1 (Haering et al., 2004). WH domains usually function in DNA binding yet no interaction between the Scc1 C-terminal region and DNA could be observed (Haering et al., 2004). It was proposed that the N-terminal domain of Scc1 could interact with Smc3 in the same way as the Smc1-Scc1 complex. However, it took over a decade to disprove this hypothesis with the solving of the crystal structure of Smc3-Scc1N, which indicated that the N-terminal region of Scc1 forms two alpha helices. These helices bind to the coiled-coil close to the head domain of Smc3 forming a helical bundle (Gligoris et al., 2014).

The additional HEAT repeat containing protein Scc3 is also required to complete the core cohesin complex (Michaelis et al., 1997). Scc3 interacts predominantly with Scc1 in an equi-molar ratio, and only partially binds to the Smc subunits (Haering et al., 2002). The crystal structure of human Scc3 and Scc1 revealed a ‘dragon’ shaped structure of Scc3. Scc1 makes a large number of contacts with the surface of Scc3 in an elongated fashion (Hara et al., 2014). A budding yeast homologue has also been solved, which predicts a similar binding mechanism between Scc3 and Scc1 (Roig et al., 2014). Furthermore, two recent crystal structures showed that a HEAT repeat protein called Pds5 binds to the N-terminal region of Scc1 through a similar mechanism to the Scc3-Scc1 interaction. This suggests that the binding mechanism of HEAT repeat proteins to subunits of cohesin is conserved (Lee et al., 2016; Muir et al., 2016).

1.1.3 The hinge domain of cohesin

Before the crystallisation of the bacterial hinge domain it was unknown whether the coiled-coil domains of Smc complexes were intra- or intermolecular. Through

structural analysis of the coiled-coils leaving the hinge domain, and EM analysis of a single Smc subunit it was clearly shown that Smc proteins have intramolecular coiled-coils (Haering et al., 2002). The crystal structure of the hinge domain also showed a pseudo 2-fold symmetry in each subunit with a positively charged channel between the dimerisation domain. The structure showed no similarity with any known protein structures. In addition, a further crystal structure of an Smc1 and Smc3 hinge complex from vertebrates showed conservation with the bacterial structure, which suggested the dimerisation mechanism has been conserved throughout evolution (Kurze et al., 2010). Biochemical analysis of the hinge domain with a region close to the coiled-coil revealed it was able to bind to double stranded DNA (dsDNA); however, the hinge domain alone was unable to (Chiu et al., 2004). The above results, raised the possibility that the hinge domain was involved in either entry or exit of DNA from cohesin's ring.

1.2 The Cohesion Cycle

The interplay between cohesin and chromatin is a dynamic process requiring many additional proteins, and is often referred to as the 'cohesion cycle'. This cycle can broadly be split into four stages: loading, anti-establishment, establishment and release (Figure 1.2). Each stage is tightly controlled throughout the cell cycle and will be reviewed in detail below. The progress towards a complete understanding of cohesin and the cohesion cycle has been the work of decades of research including the labour of many laboratories.

1.2.1 Cohesin loading

For cohesin to carry out its dynamic function it needs to be present on chromatin. This process has in recent years been shown to be complicated requiring a separate protein complex formed from two proteins, Scc2 and Scc4 (Ciosk et al., 2000). This complex was first identified in budding yeast to be required for the loading of cohesin onto DNA in the G1 phase of the cell cycle (Ciosk et al., 2000). Later, Scc2 and 4 were shown to be conserved in all higher organisms (Seitan et al., 2006). However, the bacterial cohesin equivalent does not contain a loader complex. Cohesin loading

is not just confined to G1 as studies in both budding yeast and human cells have shown that the presence of cohesin on chromatin is dynamic. For instance, Scc1 produced after the G1 phase in budding yeast was found on chromatin (Ocampo-Hafalla et al., 2007). In addition, the residence time of cohesin on chromatin in the G1 and G2 phase of the cell cycle was discovered to be ~25 minutes in HeLa cells (Gerlich et al., 2006). There appears, however, to be cohesin molecules with a much longer residence time. It was suggested that this population was responsible for the stable linkage between sister-chromatids (Gerlich et al., 2006).

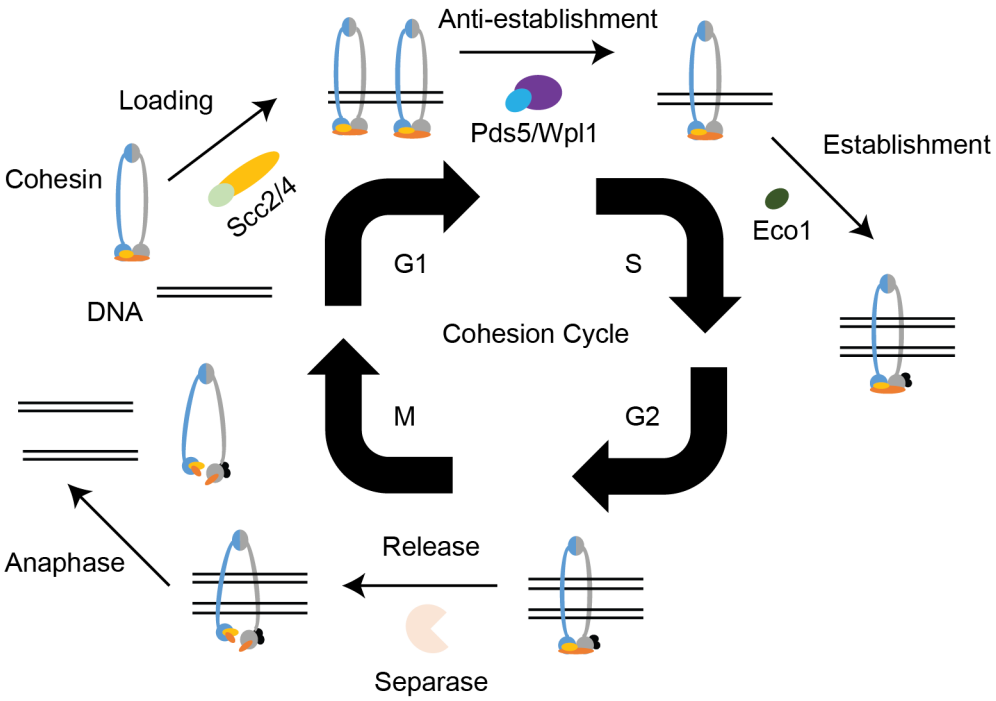


Figure 1.2. The Cohesion Cycle.

During G1, cohesin is loaded onto DNA by the Scc2/4 complex. At this stage, cohesion is known not to be established due to the anti-establishment activity of Wapl and Pds5. During S-phase cohesion is established by the acetyltransferase Eco1. Finally, upon the onset of anaphase, cohesion between sister-chromatids is lost due to the action of separase, which cleaves Scc1.

Analysis of Scc2 revealed it contained a C-terminal HEAT repeat domain (Neuwald and Hirano, 2000) with further structural studies indicating that the loader complex formed three separate domains. These globular regions were called the head, body and hook regions (Chao et al., 2015). Scc2 is also able to bind to dsDNA but not ssDNA (Murayama and Uhlmann, 2013). The hook is formed from the C-terminal

region of Scc2 and is sufficient to load cohesin onto circular DNA *in vitro* (Chao et al., 2015). This implicated Scc4 in a regulatory role for cohesin loading as it is thought to recruit Scc2 to specific chromatin locations (Bernard et al., 2006). Scc4 is a tetratricopeptide repeat (TPR) protein, which binds to an extended N-terminal peptide from Scc2, (Chao et al., 2015; Hinshaw et al., 2015) and has a conserved patch of residues on its surface required for the localisation of the loader complex (Hinshaw et al., 2015).

The Scc2/4 complex will load cohesin onto DNA, *in vitro*, in a sequence independent manner (Murayama and Uhlmann, 2013). However, in cells cohesin is loaded at actively transcribed regions before it moves to intragenic areas (Lengronne et al., 2004). This suggested that Scc4 has a role in locating the complex to specific chromatin location through interactions with a protein partner, as it has a redundant role in the loading reaction (Chao et al., 2015). This has been supported with evidence that the loader co-localises with both the mediator complex (Kagey et al., 2010) and the RSC complex (Lopez-Serra et al., 2014). In addition, Scc4 is thought to interact directly with the pre-replication complex via the Cdc7-Dbf4 subunits (Takahashi et al., 2008). It is possible that Scc4 interacts with a variety of chromatin bound complexes; however, further work needs to be conducted to clarify the above results.

Two studies carried out by Frank Uhlmann and co-workers have improved our understanding of the loading mechanism of cohesin by Scc2/4 (Murayama and Uhlmann, 2015; 2013). Using recombinant proteins, they managed to reconstitute the loading reaction *in vitro*, and provided evidence that cohesin topologically entraps DNA. They also showed that ATP hydrolysis was required for the loading reaction, which they suggested caused dissociation of the Smc head domains. Furthermore, their results indicated that DNA enters cohesin through an ‘entry’ gate, which is located between the Smc3-Scc1 interface.

1.2.2 Anti-establishment

Analysis of cohesin loaded onto chromatin has shown that it is not stably associated in the G1 phase of the cell cycle (Gerlich et al., 2006). At this stage cohesion is known

not to be established, and 'anti-establishment' is occurring due to the action of two proteins called precocious dissociation of sisters protein 5 (Pds5) and wings apart like protein (Wapl) (Hartman et al., 2000; Vernì et al., 2000). Pds5 and Wapl form a stable sub-complex *in vitro*, which is able to bind to cohesin (Murayama and Uhlmann, 2015). Co-operative binding to cohesin is required for the interaction as Pds5 or Wapl alone could not bind independently of the other subunit (Murayama and Uhlmann, 2015). Pds5 is an essential gene as Wapl is dispensable in budding yeast (Hartman et al., 2000). This suggests Pds5 has a role outside its function in anti-establishment.

Wapl was first identified in *Drosophila melanogaster* as a non-essential gene, which affected the structural organisation of heterochromatin (Vernì et al., 2000). The yeast structure is mainly α helical (Kueng et al., 2006), and is formed from a C-terminal HEAT repeat, which is able to bind to the head domain of Smc3 (Chatterjee et al., 2013). The C-terminus is conserved between species, with the structure of human and yeast Wapl being similar (Ouyang et al., 2013). Conversely, the N-terminal region of Wapl is divergent; apart from a conserved tyrosine-serine-arginine (YSR) motif (Ouyang et al., 2016) which is required for the interaction with Pds5 (Shintomi and Hirano, 2009).

Three recent structures of Pds5 have revealed it forms a 'jaw' like HEAT repeat structure able to bind to Scc1 close to the Scc1-Smc3 interaction site through a conserved patch of residues on the surface of Pds5 (Lee et al., 2016; Muir et al., 2016; Ouyang et al., 2016). This has further been supported by crosslinking data, which shows Pds5 forms multiple contacts with Smc3 and Scc1 (Huis in 't Veld et al., 2014). In one of these crystal structures the co-factor inositol hexakisphosphate (IP_6) was bound, which was shown to participate in the opening of the exit gate of cohesin (Ouyang et al., 2016).

The topological entrapment of DNA by cohesin is catalysed by Scc2/4, with DNA passing through an 'entry' gate found between the Smc3-Scc1 interface (Murayama and Uhlmann, 2013). Opposing this action is Wapl and Pds5, which cause the removal of DNA from inside cohesin's ring through a so called 'exit' gate. Much speculation and work has been conducted into understanding how DNA both enters and leaves cohesin's topological embrace. The exit gate has been characterised

both *in vitro* and *in vivo*, and work has shown that the binding of ATP is required *in vitro*, but surprisingly not its hydrolysis; suggesting the nucleotide is required to cause dimerisation of the head domain (Murayama and Uhlmann, 2015). The exit gate has been shown to be between the Smc3-Scc1 interface, which is unexpectedly in the same location as the entry gate (Chan et al., 2012). This was firstly shown by fusing Smc3 and Scc1 in yeast cells, which revealed cohesin could no longer be removed from chromatin (Chan et al., 2012). Furthermore, recent work has shown more directly that the N-terminus of Scc1 is removed from Smc3 through monitoring the degradation of an N-terminal fragment of Scc1 in wild type or Wapl deletion strains (Beckouët et al., 2016). Finally, evidence was shown, *in vitro*, that the N-terminus of Scc1 could be caused to dissociate from Smc3 by the addition of Pds5 and Wapl (Murayama and Uhlmann, 2015). Surprisingly, in this assay system Wapl and Pds5 could also act as a cohesin loader; explaining the observation that deleting Wapl in yeast cells caused a reduced amount of cohesin on chromatin (Rowland et al., 2009).

In cells, the Scc3 subunit of cohesin is required as well as Wapl and Pds5 to cause the removal of cohesin from chromatin (Shintomi and Hirano, 2009). It is interesting to note that there are stoichiometric amounts of Pds5 and the other cohesin subunits found in yeast cells; however, this is not the case for Wapl (Chan et al., 2012). It could therefore be concluded that Wapl acts like an enzyme on the Pds5-cohesin complex with Pds5 and Scc3 helping regulate and stabilise an open cohesin conformation.

If the anti-establishment activity of Wapl and Pds5 was predominant throughout the cell cycle cohesion between sister chromatids would not occur, which would result in chromosome mis-segregation. Cells have therefore evolved a mechanism to cause the controlled establishment of sister chromatid cohesion (see below), which they have coupled to DNA replication.

1.2.3 Establishment

The establishment of sister-chromatid cohesion is required to antagonise the anti-establishment activity of Wapl and Pds5. Although the protein establishment factor-

like protein 1 (Eco1) does not form part of the cohesin complex it was identified in yeast genetic screens to be an essential gene for chromosome segregation (Spencer et al., 1990; Michaelis et al., 1997). Further work revealed Eco1 was not required for maintaining sister-chromatid cohesion after S-phase or loading the complex onto chromatin. This implicated Eco1 in an establishment role, which occurred during DNA replication (Toth et al., 1999).

Eco1 is highly conserved with vertebrates containing two copies of the gene. Mutations in one of the human orthologues, Esco2, causes the two developmental disorders called Roberts syndrome and the less severe SC Phocomelia (Schüle et al., 2005). Although Eco1 is conserved the mechanism of establishment in vertebrates varies slightly from yeast where the additional protein sororin is required. Sororin aids establishment by directly competing with Wapl for Pds5's binding site (Nishiyama et al., 2010).

Early analysis on the secondary structure of Eco1 revealed it contained a two domain structure with a N-terminal C_2H_2 Zinc finger domain and a C-terminal GCN5 related N-acetyl-transferase (GNAT) domain (Ivanov et al., 2002). In addition, between species the N-terminus is divergent in both length and composition signifying species specific roles for some orthologues of Eco1. It was further shown that Eco1 possessed *in vitro* acetylation activity as it was able to auto-acetylate itself and subunits of the core cohesin complex. This suggested its main role in cells was in acetylating subunits of the cohesin complex, which would antagonise the activity of Wapl and Pds5 leading to the establishment of sister-chromatid cohesion. However, an *in vivo* target could not be identified (Ivanov et al., 2002). This caused the significance of the acetyltransferase domain of Eco1 to be questioned (Brands and Skibbens, 2005). This was further supported by the observation that mutating R222G and K222G in Eco1 caused a loss of acetylation activity *in vitro*, but yeast cells with this mutant did not show any defects in sister-chromatid cohesion (Unal et al., 2007).

Two studies provided evidence that in yeast an Eco1 mutant strain could be supressed by the Smc3 mutation K113N, which would mimic a lysine acetylation at this residue (Ben-Shahar et al., 2008; Unal et al., 2008). Analysis on Smc3 by mass spectrometry (MS) also confirmed that not only was K113 acetylated *in vivo* but the

neighbouring residue, K112, was as well. This tandem lysine acetylation of Smc3 was inhibited in yeast cells with mutant Eco1. The lysine acetylation was confirmed to be conserved as the corresponding residues in human Smc3, K105 and K106, were also found to be acetylated (Zhang et al., 2008). This provided the final evidence that Eco1 causes the establishment of sister-chromatid cohesion by acetylating Smc3 on residues K112 and K113 (human K105 and K106).

The recent crystal structure of human Eco1 (hEsco1) has revealed that the C-terminal region forms the standard GNAT family fold (Kouznetsova et al., 2016). However, it has an extended β-hairpin that protrudes from this globular domain, which is not found in other GNAT proteins. This insertion is conserved between species and is important in forming a dimer interface in solution. Furthermore, the authors show that hEsco1 is missing a conserved catalytic glutamate found in other GNAT proteins. They suggest that an aspartate from Smc3 functions in substrate-assisted catalysis (Kouznetsova et al., 2016).

Since the discovery that a post translational modification of K112 and K113 causes a change to the cohesin complex which triggered the establishment of sister-chromatid cohesion speculation has ensued about the effect acetylation has on the cohesin complex. There have been three popular ideas for the functional significance of acetylation: (1) acetylation regulates the ATPase activity of the head domain of cohesin (2) modification of Smc3 inhibits Pds5 or Wapl binding and (3) DNA binding is inhibited upon Smc3 acetylation. The evidence for the effect acetylation has on the ATPase activity of cohesin is varied with studies showing mixed results (Çamdere et al., 2015; Heidinger-Pauli et al., 2010; Ladurner et al., 2014). However, in vitro, when acetylation mimics were made recombinantly there was no change in the background ATPase activity of cohesin (Ladurner et al., 2014). In yeast cells, contradictory to this, acetylation seems to locks cohesin in a state that is unable to hydrolyse ATP (Elbatsh et al., 2016). Recent evidence has shown that loader induced ATPase activity of cohesin is inhibited by acetylation (Murayama and Uhlmann, 2015). The authors suggest that the tandem lysines act as sensors, regulating the ATPase activity of the head domain. This has been hypothesised to be due to acetylation blocking DNA binding. Evidence for this model is reinforced by the crystal structure of Rad50 bound to DNA, which shows that the loop corresponding to K112 and K113 of Smc3 is in

the proximity of DNA (Rojowska et al., 2014). Support for acetylation inhibiting Wapl binding has recently been provided, as Wapl cannot cause the dissociation of the exit gate when Smc3 is acetylated *in vivo* (Beckouët et al., 2016). However, ATPase stimulated activity could be required to cause the dissociation of the exit gate; limiting the scope of this study. Further work needs to be conducted to clarify the functional role of acetylation.

The functional significance of acetylation is an important question that needs answering. In addition to this, another vital avenue of research has been investigating the regulation of Eco1 as the timing of acetylation is crucial. Early work implicated that the establishment of sister-chromatid cohesion was linked with DNA replication (Toth et al., 1999). One model suggested that Smc3 is acetylated as the replication fork passes through cohesin. A potential problem with this model is that the replication fork may be too large to pass through cohesin's ~50nm ring. It is possible that either cohesin is unloaded and then re-loaded directly after the replication fork or the replication fork undergoes a conformation change when it encounters cohesin (Uhlmann, 2009). The idea that establishment is coupled to the passing of the replication fork was supported by the discovery that Eco1 interacts directly with proliferating cell nuclear antigen (PCNA), which is a DNA sliding clamp (Moldovan et al., 2006). PCNA is a crucial replication factor able to bind multiple proteins through a conserved PCNA interaction motif or PIP box (Moldovan et al., 2007). However, direct binding was not shown in this study, and the PIP box identified did not contain two key aromatic residues, which are usually crucial in PCNA binding (Moldovan et al., 2007).

Eco1 is also regulated by Pds5, as Pds5 acts outside its anti-establishment role to promote acetylation *in vivo* (Chan et al., 2013; Vaur et al., 2012). It has yet to be understood how Pds5 performs this role; whether it is directly through interacting with Eco1 or by stabilising the head domain of cohesin. This could explain why Pds5 is an essential gene but Wapl is not (Vernì et al., 2000). *In vitro* the hydrolysis of ATP is required for the acetylation of Smc3, suggesting Eco1 can only interact with the head domain when it is in the correct conformation (Ladurner et al., 2014). As well as being spatially regulated by PCNA, Eco1 has a cyclin-dependent kinase 1 (Cdk1)

phosphorylation site. Phosphorylation of Eco1 occurs after S-phase causing its eventual degradation (Lyons and Morgan, 2011).

Apart from Eco1, a number of other establishment factors have been discovered, which although they are non-essential cause chromosome mis-segregation defects. Currently, these non-essential factors include Ctf4 (Hanna et al., 2001), Ctf18, Dcc1, Ctf8 (Mayer et al., 2001), Tof1, Csm3, Mrc1 (Warren et al., 2004) and Chl1 (Petronczki, 2004). Of these proteins only Ctf18, Dcc1, Ctf8, Csm3, Ctf4 and Chl1 have been directly linked to cohesion establishment (Borges et al., 2013; Xu et al., 2007). The above proteins have been separated into two distinct groups from their genetic interactions. The first group is made from Mrc1, Ctf18, Dcc1 and Ctf8; and the second group contains Tof1, Csm3, Ctf4, Chl1.

Ctf18 forms a replication factor C (RFC^{Ctf18}) like complex with the proteins Dcc1 and Ctf8 (Mayer et al., 2001). This complex is present at replication forks during S-phase (Lengronne et al., 2006). Apart from sister chromatid cohesion, RFC^{Ctf18} also functions with Mrc1 in intra S-phase checkpoints (Crabbé et al., 2010). Ctf4 forms a trimer able to interact with multiple components of the replisome (Simon et al., 2014). Recently, it has also been shown to interact directly with Chl1 (Samora et al., 2016). Chl1 is a DNA helicase (Hirota and Lahti, 2000) first identified in chromosome mis-segregation screens. Recently, it has been implicated in lagging strand replication (Farina et al., 2008). Chl1 and Ctf4 function in an Eco1 independent establishment pathway by a currently unknown mechanism (Borges et al., 2013).

Further work needs to be conducted into Eco1 and the role of acetylation, as well as the other non-essential establishment factors for a more complete understanding of this crucial stage in the cohesion cycle. Genetic work alone will be unable to unearth a complete functional model, and further biochemical and structural analysis will be required.

1.2.4 Release of cohesin

After the establishment of sister-chromatid cohesion, newly replicated sisters remain topologically entrapped in cohesin's ring until the onset of anaphase when cohesion is lost (Nasmyth and Haering, 2009). Without this physical linkage the positioning of chromosomes on the metaphase plate would be an almost impossible task for cells. Early work in budding yeast implicated the cysteine protein separase in the removal of cohesin from chromatin (Uhlmann et al., 2000). Furthermore, it was shown, *in vitro*, that separase could cleave the subunit Scc1 via two conserved protease sites found between the N- and C- terminal Smc binding regions (Hornig et al., 2002). Apart from Scc1, separase is also able to cleave other proteins involved in promoting anaphase, and its function seems to be conserved in all organisms (Hauf et al., 2001).

Separase is large protein, which is over 2000 residues long in many species. From secondary structure prediction it was shown to have an N-terminus made from 26 Armadillo (ARM) repeats followed by a central region, which was predicted to be unstructured (Hornig et al., 2002). After this central region, separase contains two protease domains, of which one is non-functional. These domains are cysteine proteases belonging to a family that comprises caspases (Hornig et al., 2002).

The early release of sister-chromatids from cohesin would be disastrous for cells, leading to chromosome mis-segregation. Therefore, like the other stages of the cohesion cycle, the release of cohesin is tightly controlled. There are two known mechanisms to regulate separase (Gorr et al., 2005; Uhlmann et al., 2000). The first method involves the inhibitory chaperone called securin, which binds to the protease domain of separase acting as an inhibitory substrate (Uhlmann et al., 2000). In addition, Cyclin B is able to bind and inhibit separase once separase is phosphorylated by Cdk1 on residue S1121 (Gorr et al., 2005). On the onset of anaphase separase becomes active as APC ubiquitinates both cyclin B and securin. This leads to their eventual degradation by the proteasome, which allows separase to become fully active (Uhlmann et al., 2000).

Once Scc1 is cleaved by separase sister-chromatids can be pulled to opposite poles of the cell. The cleavage of Scc1 means it needs to be replaced in the next cell cycle;

however, the other core subunits are reused. This means Smc3 must become deacetylated during anaphase to complete the cohesion cycle as acetylated cohesin cannot be loaded onto DNA (Borges et al., 2010). The Class 1 histone deacetylase called Hos1 has been implicated as the cohesin deacetylase in budding yeast (Borges et al., 2010). It has been shown that Hos1 becomes activated after cohesin is removed from chromatin, and Hos1 is only able to deacetylate cohesin once Scc1 has been cleaved (Borges et al., 2010).

1.2.5 The prophase pathway

Although separase cleavage is the main contributor to the removal of cohesin from chromatin in yeast, higher eukaryotes have an additional mechanism to dissociate cohesin from DNA. This occurs exclusively during prometaphase and prophase, and is termed the 'prophase pathway' (Sumara et al., 2000). During prophase, in higher eukaryotes, cohesion is lost at chromosome arms but is maintained at the centromere region (Sumara et al., 2000). During mitosis, in budding yeast, cohesin is not removed during prophase but the prophase pathway does occur during meiosis. This suggests that the higher eukaryotic mitotic prophase pathway and yeast meiotic prophase pathway both have a common ancestor (Kitajima et al., 2004).

Separase is not active during prophase; therefore, the loss of cohesin from chromatin during the prophase pathway occurs via a different mechanism (Sumara et al., 2000). To activate the removal of cohesin during prophase two kinases are required to phosphorylate the C-terminal domain of Scc3 called Aurora B and a polo-like kinase (PLK) (Hauf et al., 2005; Sumara et al., 2002). In *X. laevis* cells, either a reduction of PLK or Scc3 phosphorylation mutations reduces the activity of this pathway (Waizenegger et al., 2000). Aurora B did not act as a kinase suggesting it does not phosphorylate cohesin during prophase (Losada et al., 2002). To aid in the recruitment of PLK, Cdk1 has been shown to phosphorylate Sororin at T159. Mutating this residues meant PLK was not recruited to cohesin, and this caused a disruption in the prophase pathway (Zhang et al., 2011). In addition, the anti-establishment factor Wapl is crucial for the removal of cohesin from chromosome

arms during prophase as mammalian cells depleted of Wapl caused cohesin to remain on arm regions during prophase (Gandhi et al., 2006).

If cohesin was completely removed from chromatin, then sister-chromatid cohesion would be lost and chromosome mis-segregation would occur. To protect a proportion of cohesin, centromeric cohesin is maintained. The two proteins shugoshin 1 (Sgo1) and protein phosphatase 2A (PP2A) have been found to antagonise the action of PLK (Kitajima et al., 2006). Depletion in human cells of Sgo1 resulted in chromosome mis-segregation (Salic et al., 2004), and Sgo1 knock out mice are unviable (Lee et al., 2008). Both results show the importance of this protein in maintaining sister-chromatid cohesion at the centromere. In addition to protecting cohesin at the centromere, Sgo1 has been found in small quantities at the arms of chromosomes, which suggested a proportion of cohesin is protected from Wapl removal during prophase (Shintomi and Hirano, 2009). Sgo1 acts in a complex with PP2A, as work revealed they interacted *in vitro* (Kitajima et al., 2006). Both proteins act together to de-phosphorylate Scc3, which protects cohesin from Wapl induced removal (Kitajima et al., 2006). Further pull down experiments showed that Sgo1 directly competes with Wapl for its binding site on Scc3 (Hara et al., 2014). In addition, Sororin is required to be kept in a hyper-phosphorylated state to help antagonise the activity of Wapl at the centromere. Unexpectedly, Sgo1 and PP2A have been shown to help with this process (Liu et al., 2012).

From the above discussion on the cohesion cycle it can be clearly seen that each stage is tightly regulated, and requires the co-ordination of numerous proteins. In the next section I will discuss in more detail the RFC^{Ctf18}complex, which is a non-essential establishment factor.

1.3 The Function of RFC^{Ctf18}

The RFC^{Ctf18} complex was first discovered in budding yeast to cause chromosome mis-segregation in genetic screens (Mayer et al., 2001). Further work showed that it was directly involved in the establishment of sister-chromatid cohesion, as deletion of Ctf18 caused a reduction in acetylation of Smc3 by Eco1 (Borges et al., 2013). In

this section, work relating to RFC^{Ctf18} is reviewed in detail, along with the discovery that it is involved in the DNA damage response checkpoint (Crabbé et al., 2010).

1.3.1 The function of RFC^{Ctf18} in cohesion establishment

Early work indicated that Ctf18 was not essential but was important for chromosome segregation (Kouprina et al., 1993). A later study suggested that Ctf18 had a sequence which was similar to Rfc1, which indicated it could have an equivalent role (Kouprina et al., 1994). Rfc1 forms a 5 subunit complex with Rfc2-5, which is called RFC^{Rfc1}. RFC^{Rfc1} functions during replication, and is able to load PCNA on both the lagging and leading strand (Kubota et al., 2014). Further work revealed that Ctf18 interacts both physically and genetically with proteins known to function at the replication fork (Hanna et al., 2001). In this study using an *in vivo* cohesion assay, which observed the early separation of sister-chromatids, they also suggested that Ctf18s role in chromosome stability was due to its function in sister-chromatid cohesion (Hanna et al., 2001). Furthermore, chromatin immunoprecipitation (ChIP) experiments revealed that Ctf18 co-localised on chromatin with the replication fork and other establishment factors such as Eco1 (Lengronne et al., 2006). Monitoring the acetylation levels has shown that yeast cells without Ctf18 have a reduced level of Smc3 acetylation; suggesting Ctf18 functions to aid Eco1 acetylation of cohesin. This was further supported by the observation that Eco1 and Ctf18 interact in pull down experiments conducted in yeast cells (Kenna and Skibbens, 2003).

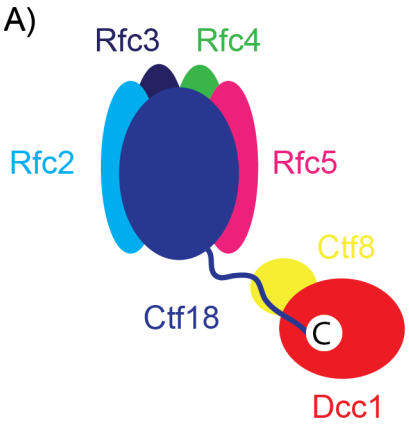
Sequence alignments originally identified Ctf18 as a homolog to Rfc1 (Kouprina et al., 1994). However, confirmation of this did not come until pull down experiments identified that Ctf18 physically interacted with the four proteins Rfc2-5; forming an RFC like complex called RFC^{Ctf18} (Mayer et al., 2004). EM analysis of this five subunit complex revealed it was structurally similar to RFC^{Rfc1}, and like RFC^{Rfc1}; RFC^{Ct18} was able to interact with PCNA and load the trimer onto DNA *in vitro* (Shiomi et al., 2004). This was supported by the observation with ChIP analysis that deleting Ctf18 in yeast caused a decrease in the amount of PCNA present on chromatin (Lengronne et al., 2006). However, a more complete *in vitro* study using recombinant proteins, indicated that RFC^{Ctf18} could both unload and load PCNA onto a range of DNA

substrates (Bylund and Burgers, 2005). Furthermore, this work suggested that, *in vivo*, under physiological condition RFC^{Ctf18} would act as a PCNA unloader. Taken together these results imply that RFC^{Ctf18} may act both as a loader and unloader of PCNA from chromatin depending on the requirement of the cell to shift the equilibrium of this reaction. This complex mechanism will need further work to clarify how cells are able to regulated the interaction between PCNA and RFC^{Ctf18}.

In addition to binding to Rfc2-5, Ctf18 also interacts with the stable sub-complex formed from defective in sister-chromatid cohesion protein 1 (Dcc1) and chromosome transmission fidelity protein 8 (Ctf8) (Mayer et al., 2001). This sub-complex interacts both genetically and physically with Ctf18. Both Ctf8 and Dcc1 have been shown to be non-essential cohesion establishment factors, as mutations in either protein causes an increase rate of chromosome mis-segregation in yeast (Mayer et al., 2001). Furthermore, both Ctf8 and Dcc1 are conserved throughout evolution with human Dcc1 being overexpressed in colorectal cancer (Yamaguchi et al., 2014). Pull down experiments revealed that the Dcc1-Ctf8 sub-complex interacts tightly with Ctf18 and Rfc2-5 (Mayer et al., 2001). This indicated that RFC^{Ctf18} is a 7 subunit complex formed from Ctf18, Rfc2, Rfc3, Rfc4, Rfc5, Dcc1 and Ctf8 (Figure 1.3A). Interestingly, the PCNA loading or unloading activity of RFC^{Ctf18} was not inhibited by removing Dcc1 or Ctf8. This suggested that the Dcc1-Ctf8 sub-complex had a function in regulating the RFC^{Ctf18} complex (Bylund and Burgers, 2005). A study with recombinant proteins showed that the C-terminal region of Ctf18 was highly conserved, and a deletion of the last 80 residues of Ctf18 was enough to disrupt binding to the Dcc1-Ctf8 sub-complex (Figure 1.3B) (Bylund and Burgers, 2005). Further work showed that conserved aromatic residues present in the C-terminal domain of Ctf18 was responsible for the interaction with Dcc1 and Ctf8 (García-Rodríguez et al., 2015). Furthermore, mutating these aromatic residues to alanine, in yeast, caused an increase in chromosome mis-segregation. The above results suggested that all seven subunits are required for RFC^{Ctf18} to perform its role in sister-chromatid cohesion.

To improve our understanding of RFC^{Ctf18} further work will need to be conducted into the two accessory subunits, Dcc1 and Ctf8. These proteins will be crucial in unlocking how RFC^{Ctf18} functions in sister-chromatid cohesion as they appear to be

the key regulators of this complex. In the next section RFC complexes are further reviewed, looking more closely at the variety of tasks they perform throughout the cell cycle.



B) Ctf18 C-terminal

```
S.cerevisiae    LEETQKTIGSDETNQADDCNQTVKIWVKYNEGFSNAVRKNVTWNNLWE- 741
S.pombe         SEAVTSN-NAALNTGDHPVSIKHAINIKFHDGFSNAVRKPISLNEILNF 960
X.laevis        TA-SANQEESVE-RRIGKAVGNSDVWFRFNEGVSNAVRRNIYIKDLL-- 1000
M.musculus      REVEAPQKDADE-WRMGVAVGRSEVWFRFNEGVSNAVRRSLYIRDLL-- 969
H.sapiens       AGDTAPEQDSVE-RRMGTAVGRSEVWFRFNEGVSNAVRRSLYIRDLL-- 975
                          :              : .:::.:*.*****: :  .::
```

Figure 1.3. Organisation of RFC^{Ctf18}.

A) Diagram of RFC^{Ctf18}. Ctf18 (dark blue) has an extended C-terminal domain that binds to the Ctf8 (yellow)-Dcc1 (red) sub-complex. B) Sequence alignment of the conserved C-terminal region of Ctf18. The highlighted red residues indicate a conserved SNAVR motif, and the highlighted yellow residues show hydrophobic residues important for interacting with Dcc1 and Ctf8. The stars (*), dots (.) and colons (:) represent positions with fully conserved amino acids, positions with comparable properties or positions with weakly similar properties respectively.

1.3.2 Comparison of RFC like complexes

RFC like complexes are formed from at least five subunits. They all contain the four small subunits Rfc2, Rfc3, Rfc4 and Rfc5; and one larger subunit, which defines the function of the complex (Figure 1.4A). Each subunit belongs to the family of proteins called AAA+ ATPases, which show sequence conservation (Kubota et al., 2014). In

addition, all of the complexes interact with DNA clamps either loading or unloading them from chromatin.

There are 4 RFC complexes encoded in the eukaryotic genome; defined by the large subunit (Figure 1.4B). RFC^{Rfc1} is the prototypical member, and is one of the most well studied RFC complexes. It has been shown to load PCNA at the start of DNA replication on both the lagging and leading strand of DNA replication (Hedglin et al., 2013). PCNA is a trimer, forming a ring shaped complex that once loaded, topologically entraps DNA (Moldovan et al., 2007). It operates in replication as a polymerase clamp, and acts as a platform for additional proteins to be recruited. To load PCNA onto DNA, RFC^{Rfc1} requires ATP binding to open the ring of PCNA. It then loads PCNA at a 3' primer template junction with ATP hydrolysis causing the dissociation of RFC^{Rfc1} from PCNA (Hedglin et al., 2013). The only complex that interacts with a different DNA clamp is RFC^{Rad24}. This complex loads the Rad9-Hus1-Rad1 (9-1-1) clamp at sites of DNA damage, which activates a DNA damage checkpoint response by recruiting factors that interact specifically with this clamp (Green et al., 2000). Until recently, RFC^{Elg1} was the least understood clamp loader complex as, *in vitro*, data indicated it could bind to PCNA but was unable to load or unload it from DNA. However, recent evidence has suggested it is vital for unloading PCNA at the termination of replication (Kubota et al., 2013). This would explain the early genetic work, which showed that Elg1 was important for chromosome stability and maintenance. The final clamp loader is RFC^{Ctf18}, which is the only complex formed from seven subunits. As discussed above, this complex functions in the establishment of sister-chromatid cohesion.

In addition to the function of RFC^{Ctf18} in the establishment of sister-chromatid cohesion, more recent work has implicated RFC^{Ctf18} in the DNA replication checkpoint (DRC) (Crabbé et al., 2010). The authors used ChIP analysis to show that the firing of late origins was inhibited by Ctf18 under replication stress. This revealed that RFC^{Ctf18} was involved in the DRC as under replication stress the firing of late origins was delayed. Furthermore, the deletion of Ctf18 resulted in a reduction of phosphorylation of Rad53; a read-out for the activation of this checkpoint (Crabbé et al., 2010). Three studies have suggested that RFC^{Ctf18} carries out its role in the DRC through a direct interaction with polymerases-ε (García-Rodríguez et al., 2015;

Murakami et al., 2010; Okimoto et al., 2016). Pull down and yeast two-hybrid assays were conducted in human and yeast cells respectively. The results showed that RFC^{Ctf18} directly bound to polymerases-ε (García-Rodríguez et al., 2015; Murakami et al., 2010). This was suggested to occur through the heterotrimer formed from Ctf18-Dcc1-Ctf8, which created a binding module for the N-terminal region of Pol2, which is a subunit of polymerases-ε (García-Rodríguez et al., 2015).

RFC complexes are key regulators of the cell cycle helping preserve genome integrity. In the next section intra-S checkpoints will be reviewed, with analysis of the DRC and DNA damage checkpoint (DDC).

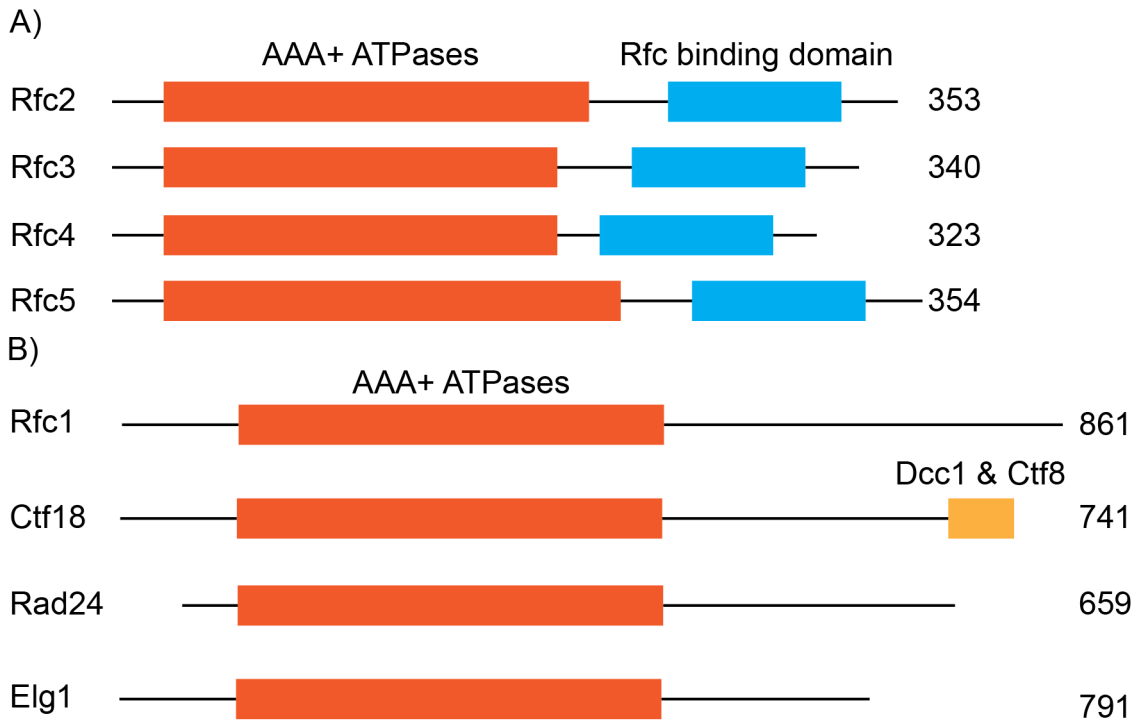


Figure 1.4. Subunits of the RFC Like Complexes.

A) The domain organisation of Rfc2, Rfc3, Rfc3 and Rfc5, with the AAA+ ATPase domain in orange and the region responsible for complex formation in blue. B) The large subunits defining the RFC complexes. AAA+ ATPase domain are indicated in orange. Ctf18 has the additional Dcc1-Ctf8 binding domain present at its C-terminus (brown).

1.3.3 Intra-S phase checkpoints

To maintain genome integrity cells have evolved checkpoints, which act as pause points during the cell cycle. In budding yeast, checkpoints exist between G2/M and G1/S transitions. In addition, there are intra-S checkpoints that function to provide cells time to repair damaged DNA during replication stress. The intra-S checkpoint causes a range of downstream effects such as the suppression of origin firing, expression of DNA stress genes, replication fork stabilisation and inhibiting cell cycle transition (Branzei and Foiani, 2009). Two phosphoinositide 3-kinases (PI3K) are crucial in intra-S checkpoint response, called Mec1 and Tel1 (ATR and ATM respectively in humans). These two kinases respond to different stimuli: Mec1 is activated by base adducts, replication stress, ultra-violet (UV) induced damage and double strand breaks (DSBs); as Tel1 responds primarily to DSBs (Cimprich and Cortez, 2008). In addition, Mec1 binds to the protein Dcd2 (ATRIP in humans) and this interaction is required to activate it (Majka et al., 2006).

Analysis of the Mec1 checkpoint by a genome wide study in *Saccharomyces cerevisiae* revealed that this pathway could further be separated into two branches based on their genetic interactions, called the DDC and DRC (Pan et al., 2006). Furthermore, later work supported this conclusion as the authors showed that they could isolate the DRC and DDC based on their role in the suppression of the firing of late replication origins (Crabbé et al., 2010). In this study, ChIP analysis was conducted using BrdU incorporation as an analysis of replication origin firing. Both pathways require the activation of Mec1. However, it is the way in which Mec1 becomes activated that determines the downstream effects (Crabbé et al., 2010). In both the DDC and DRC Rad53 (Chk2 in humans) becomes activated by Mec1 as without this kinase Rad53 becomes hyper-phosphorylated at a slower rate (Crabbé et al., 2010).

The activation of the DRC (Figure 1.5) occurs in budding yeast by an accumulation of ssDNA (Aparicio et al., 1999). However, a recent study using *X. laevis* egg extracts indicated that primed ssDNA was required to activate the DRC along with Polα/primase (Cimprich and Cortez, 2008). In cells, exposed ssDNA is prone to breakages or the formation of secondary structure elements. Therefore, a protein

called replication protein A (RPA) has evolved to bind and stabilise ssDNA. RPA is composed of three subunits with oligonucleotide binding motifs: RPA1, RPA2 and RPA3. The build-up of RPA-coated ssDNA is detected by Mec1, which then leads to Rad53 hyper-phosphorylation (Aparicio et al., 1999). The activation of Rad53 is mediated by the protein Mrc1. As well as mediating Rad53 hyper-phosphorylation, Mrc1 (Claspin in humans) is thought to stabilise forks through binding to Polε through its C-terminal domain (Lou et al., 2008). In addition to stabilising stalled forks, Rad53 aids in keeping the polymerase localised at the replisome (Cobb et al., 2003). As discussed above, RFC^{Ctf18} has been implicated in assisting in the activation of the Mrc1-Rad53 checkpoint. Deleting either Ctf18, Dcc1 or Ctf8 causes both a reduction and delay in the hyper-phosphorylation of Rad53 in Hydroxyurea (HU) (Crabbé et al., 2010). HU reduces the pool of dNTPs, which causes replication stress and hence the stalling of replication forks. This then leads to the activation of intra-S checkpoints and the eventual hyper-phosphorylation of Rad53. Therefore, monitoring Rad53 phosphorylation in cells grown in HU can help clarify proteins involved in stimulating the DRC or DDC. The results suggest, along with Mrc1, RFC^{Ctf18} must in some way aid in the activation of Rad53 by Mec1. It has been proposed that RFC^{Ctf18} carries out this function by unloading or loading PCNA from stalled replication forks, which could then recruit Mec1. However, further work needs to be conducted to uncover, with more clarity, the mechanism in which RFC^{Ctf18} functions. The use of Mrc1 separates the function of the DRC from the DDC, as the DDC uses the protein Rad9 to activate Rad53 (Pan et al., 2006).

The DDC uses similar proteins to the DRC; however, it becomes activated predominantly through DNA breaks. In addition, it stimulates a different branch of the Mec1-Rad53 checkpoint (Putnam et al., 2009). Although the DDC and DRC are comparable, their relative kinetics are different. The DDC is slower taking minutes to become activated. However, the DRC is fast causing a response in seconds, but this checkpoint requires numerous forks to be stalled before it is stimulated (Shimada et al., 2002). In the DDC, the Rad24 variant of the RFC like complexes loads the 9-1-1 clamp at sites of DNA damage (Navadgi-Patil and Burgers, 2008). The 9-1-1 clamp can recruit other DNA damage proteins, which aids in the checkpoint response. Rad9 (53BP1 in humans), like Mrc1, is then able to mediate Rad53 hyper-phosphorylation by Mec1 (Putnam et al., 2009).

As can be seen from the above discussion, these checkpoints use multiple proteins to monitor DNA replication. It is interesting that both the DRC and DDC use RFC like complexes to help in the stimulation of Rad53. Further work will be needed to uncover how the cohesion RFC functions in the DRC.

In the final section of this chapter work is reviewed relating to the range of diseases that have been uncovered to be caused by mutations in the cohesin specific subunits. Special attention is given to Eco1, which causes the rare genetic condition called Roberts syndrome.

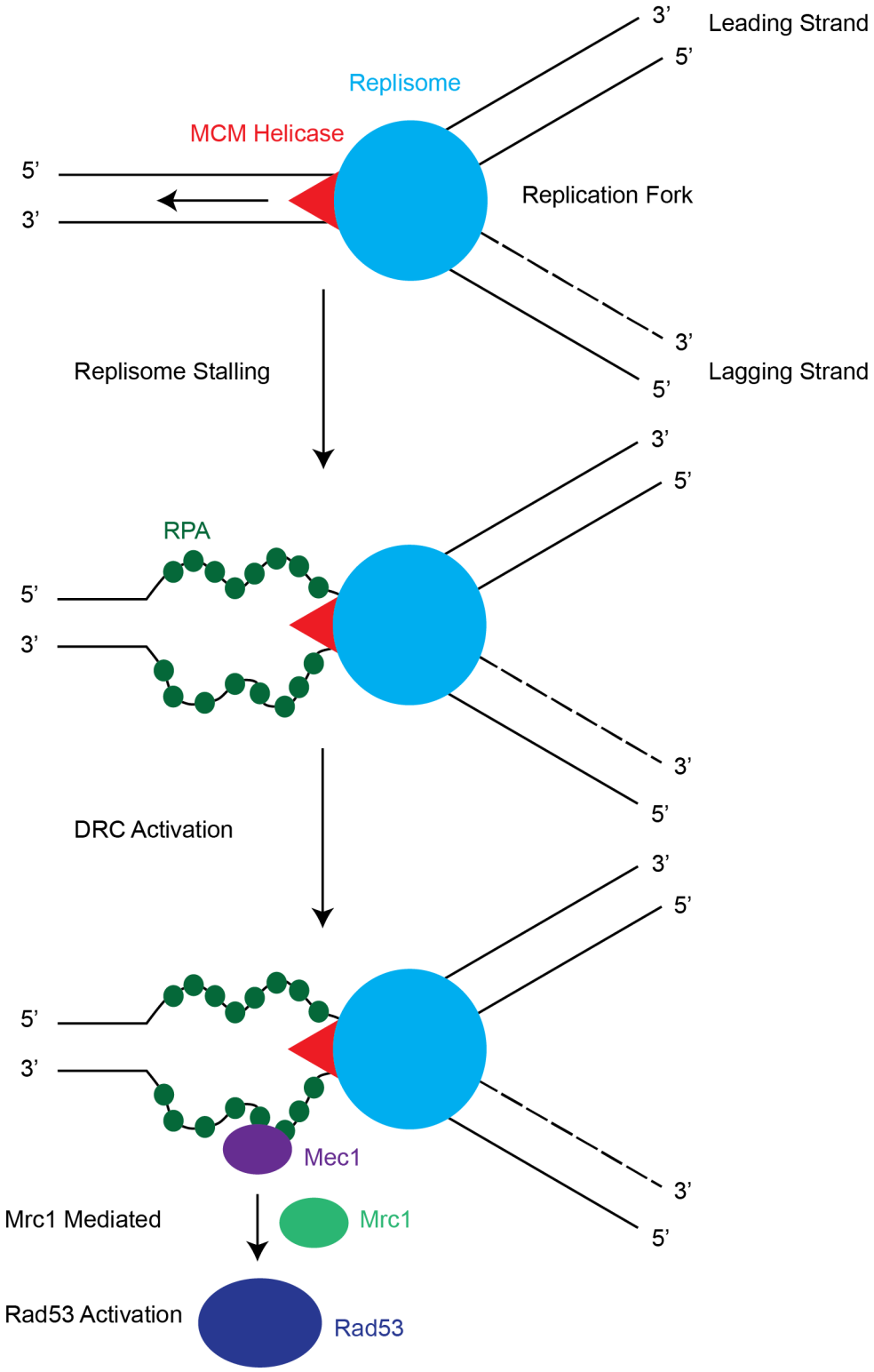


Figure 1.5. Proposed Activation Mechanism of the DRC.

During replication stress the replisome can stall. This leads to a build-up of RPA coated ssDNA, which Mec1 recognises. Mrc1 can then mediate the hyper-phosphorylation of Rad53 by Mec1.

1.4 Role of Cohesin in Diseases

The cohesin complex performs a number of functions outside its role is sister-chromatid cohesion. For instance, during interphase cohesin binds to DNA to form chromatin loops (Remeseiro et al., 2013b) and cohesin has been shown to directly affect transcription (Remeseiro et al., 2013a). It is therefore unsurprising, due to the variety of tasks performed by cohesin, that mutations in its subunits cause diseases. Not only has cohesin recently been linked to cancers, there are also a range of genetic conditions called cohesinopathies, which have been shown to be due to mutations in cohesin specific subunits. In this section, cohesins role in cancers and the two genetic disorders Cornelia de Lange syndrome (CdLS) and Roberts syndrome are discussed.

Although the specific role of cohesin in cancers is still poorly understood it is known that a number of cancers such as bladder and acute myeloid leukaemia have been shown to contain mutations in cohesin subunits (Barber et al., 2008). The recent analysis of 4742 cancer samples has revealed that SA2 (human Scc3) is mutated in a significant number of tumours (Solomon et al., 2013). In addition, bladder cancers have the highest rate of mutations in SA2, with aggressive and low grade tumours having a mutation rate of 10-15% and 30% respectively. In addition, Nibl (human Scc2) has been shown to be mutated at an increased frequency in colorectal cancer (Barber et al., 2008). This was also revealed to by the case for Dcc1, which was overexpressed in this type of cancer (Yamaguchi et al., 2014). Dcc1 forms part of the RFC^{Ctf18} complex, and it will be interesting to observe if the other subunits have a similar function in cancers. Patients with high expression levels of Dcc1 have a reduced survival rate indicating the importance in improving our understanding of this complex.

Apart from having an effect in cancer progression, mutations in cohesin specific subunits cause the genetic disease called CdLS, which was first identified by Cornelia de Lange in 1933. This disease is an inherited genetically dominant developmental disorder, which effects multiple tissue types, and can be classified due to an underdeveloped nervous system and growth retardation. (Ireland et al., 1993).

The analysis of patients with CdLS revealed that it is caused by mutations predominately in three proteins Nipbl (Scc2 homologue), Smc1 and Smc3. Nipbl is the catalytic subunit of the cohesin loader complex, as Smc1 and Smc3 form an Smc heterodimer in the core cohesin complex. Analysis of patients has revealed that ~50% of patients have mutations in the gene encoding Nipbl. These mutation lead to a non-functional protein, and are either insertions or point mutations usually found in the coding regions of the gene (Gillis et al., 2004).

In addition to CdLS, the two autosomal receive disorders called Roberts syndrome and SC Phocomelia are caused by mutations in human Eoc2 (hEsco2). hEsco2 is an acetyltransferase involved in establishing sister-chromatid cohesion during S-phase. John Roberts first identified Roberts syndrome in two children. Although both Roberts syndrome and SC Phocomelia are similar they can be separated due to different facial features. They are both extremely rare with under 100 reported cases characterised as most patients survive no further than the neonatal period. Both conditions show similar phenotypes such as server mental retardation; and prenatal and postnatal growth defects (Liu and Krantz, 2008). At a cellular level, Roberts syndrome and SC Phocomelia are diagnosed due to chromosome abnormalities (Judge, 1973).

Currently 28 genetic mutations have been identified, which cause Roberts syndrome and SC Phocomelia. These range from frame-shifts, missense and point mutations (Gordillo et al., 2008). W539G and G581R are currently the only two point mutations identified (Vega et al., 2010). Both of these mutations inhibit the auto-acetylation of hEsco2 *in vitro*; suggesting, it is the loss of a functional C-terminal GNAT domain that causes Roberts syndrome. Mapping both these residues onto the hEsco1 structure reveals that W539 is buried and stabilises an α-helix close to the active site. However, G581 is close to the CoA binding site. Mutating this residue to arginine would restrict access of CoA to its binding pocket (Kouznetsova et al., 2016).

From the above discussion it can be seen that mutations in a number of cohesin specific subunits cause cancers and a range of rare genetic diseases. Further work will be needed to clarify cohesins function in these conditions. In the next sections the aims of this project and a summary of results are presented.

1.5 Aims of Project and Summary of Results

In summary, the cohesin complex is formed from the four proteins: Smc1, Smc3, Scc1 and Scc3 (Nasmyth and Haering, 2009). A variety of complexes interact with cohesin throughout the cell cycle to load, establish and release cohesin from chromatin. Mutations in cohesins accessory factors have been shown to cause cancers and diseases (Liu and Krantz, 2008). Therefore, the study of these proteins is crucial to help improve our understanding of not only their role in the cohesion cycle but also how they function in human diseases.

The establishment of sister-chromatid cohesion is carried out predominately by Eco1, a GNAT family acetyltransferase (Ivanov et al., 2002). Eco1 acetylates Smc3 on residue K112 and K113, which antagonises the action of Wapl and Pds5 (Ben-Shahar et al., 2008). The anti-establishment complex acts to eject cohesin from chromatin through an exit gate, which is positioned between the Smc3-Scc1N interface (Murayama and Uhlmann, 2013). In addition to Eco1, a range of other non-essential establishment factors have a role in stabilising cohesin bound to chromatin. These include the RFC^{Ctf18} like complex, which is the only RFC complex formed of seven subunits (Mayer et al., 2004). The additional subunits, Dcc1 and Ctf8, form a stable sub-complex, and currently, have an unknown function as they do not participate in PCNA unloading or loading (Bylund and Burgers, 2005). In addition to functioning in the establishment of sister-chromatid cohesion, RFC^{Ctf18} has been shown to participate in the DRC, which is an intra-S checkpoint (Pan et al., 2006). This checkpoint has a range of downstream effects, one of which is the stabilisation of stalled replication forks.

The aim of this thesis was to gain further insights into the range of establishment factors involved in sister-chromatid cohesion. To do this, a structural approach was taken, using the technique of x-ray crystallography alongside biochemistry.

Work was initially started on the Dcc1-Ctf8 sub-complex bound to the C-terminal region of Ctf18. I was able to crystallise and solve the atomic resolution structure of this complex. To arrive at the final structure, multiple rounds of construct design and crystal optimisation was required. The crystal structure indicated that Ctf18 bound to

the Dcc1-Ctf8 sub-complex as an extended peptide. Unexpectedly, it was shown that Dcc1 contained a C-terminal region formed of three winged helix (WH) domains, which could bind to both ssDNA and dsDNA. In addition, these WH domains supported binding to Pol2, which is a subunit of polymerase-ε. To uncover a biological function for these WH domains, ChIP-qPCR was conducted, which indicated that they were responsible for recruiting the RFC^{Ctf18} complex to chromatin.

The crystal structure of the xEco2, the *X. laevis* homologue to Eco1, was also solved with two Smc3-Acetyl CoA conjugate peptide (Ac-CoA-K105 and Ac-CoA-K106). The structure of xEco2 bound to Ac-CoA-K105 and Ac-CoA-K106 revealed that xEco2 binds to the two tandem lysine residues in different conformations. Furthermore, with the use of MS analysis I was able to show that acetylation of K105 occurs faster than the acetylation of K106.

Chapter 2. Materials and Methods

2.1 Cloning

2.1.1 Expression cloning

The general cloning was conducted by restriction free (RF) cloning (van den Ent and Löwe, 2006). Primers were synthesised (SIGMA) to match the coding region of the gene of interest with the addition of an overhang corresponding to the vector. For RF primer sequences used in this thesis see Appendix 1. A polymerase chain reaction (PCR) was carried out in Hi-Fi Buffer (Bioline), 1 mM dNTP, 0.4 mM both forward and reverse primers, 50 $mU\mu L^{-1}$ Velocity polymerase (Bioline) and 0.1 $ng\mu L^{-1}$ template DNA. The reaction was carried out at the temperatures indicated in Table 2.1. The completed PCR was purified by 0.8% agarose gel electrophoresis run at 80 V for 45 minutes. The gel was visualised by ethidium bromide exposed with ultra violet (UV) light. The PCR product was cut using a sterilised razor and gel extracted using a commercial kit (QIAGEN). A second PCR reaction was conducted in Phusion reaction buffer (NEB), 0.5 mM dNTP, 0.8 $ng\mu L^{-1}$ vector, 50 $mU\mu L^{-1}$ Phusion polymerase (NEB) and at least a 20 times molar excess of the insert compared to the vector. The second PCR was completed at the temperatures shown in Table 2.2. The reaction was subjected to DpnI (NEB) digestion for 2 hours at 37 °C. This mixture (2 µL) was transformed into chemically competent XL-1 blue cells (Agilent). Antibiotic selection was used to determine colonies which had obtained the appropriate vector. Colonies were picked and grown at 37 °C for 16 hours in the selection antibiotic. The plasmid was purified using a commercial available kit (QIAGEN) and sequenced using a standard protocol to confirm that the construct had been inserted into the vector.

	Temperature (°C)	Time (min.)	
Initial Denature	98	0.5	
Denature	98	0.5	25 cycles
Anneal	55	1	
Extension	72	0.5 min per Kb	
Final extension	72	0.5	

Table 2.1. Reaction Conditions for the First PCR in RF Cloning.

	Temperature (°C)	Time (min.)	
Initial Denature	98	0.5	
Denature	98	0.5	18 cycles
Anneal	55	1	
Extension	72	2 min per Kb	
Final extension	72	0.5	

Table 2.2. Reaction Conditions for the Second PCR in RF Cloning.

2.1.2 Mutagenesis cloning

Mutagenic PCR was conducted using the Quick-change II protocol (Agilent). Primers which contained the appropriate mutation were commercially synthesised (SIGMA). For mutagenic primer sequences used in this thesis see Appendix 2. A PCR was carried out in Pfu buffer (Agilent), 0.5 mM dNTP, 0.4 mM both forward and reverse primers, 50 $mU\mu L^{-1}$ Pfu polymerase (Agilent) and 0.1 $ng\mu L^{-1}$ template DNA. The reaction was completed at the temperatures presented in Table 2.2. The PCR was digested for 2 hours at 37 °C by DpnI. This mixture was then directly transformed into chemically competent XL-1 blue cells (Agilent) using a standard head shock protocol. Colonies were picked from an antibiotic selection plate and grown at 37 °C for 16 hours. The plasmid was purified using a commercial available kit (QIAGEN), and sequenced using a standard protocol to confirm that the point mutation had been incorporated into the gene.

2.2 Protein Expression

The constructs used in this thesis were cloned from genomic DNA into either pET-28a, pCDFDuet or pETDuet based vectors as described in section 2.1. Appropriate subunits were tagged with an N-terminal His_6 and TEV protease site. For protein expression cloned constructs were freshly transformed into BL21(DE3)-RIL competent cells (Agilent) using a standard heat shock protocol. Cells were selected with an appropriate antibiotic grown for 18 hours at 37 °C. Colonies were picked and grown in 5 mL fresh lysongeny broth (LB) for 18 hours at 37 °C in the appropriate

selection antibiotic. The overnight culture was used to inoculate 1 L of LB. Cells were grown at 37 °C with shaking at 200 rpm until an OD of 0.6 was reached. At this point, IPTG was added to induce protein expression (0.25 mM). Cultures were cooled to 16 °C and left for 18 hours; after which, cells were harvested at 6000 xg, 4 °C for 20 minutes. Pellets were washed in cold double distilled water, and snap frozen in liquid nitrogen. Frozen cells were stored for future use at -80 °C.

To obtain selenomethionine labelled protein cells were grown in the same conditions. However, the 1 L inoculated cultures were expanded in minimal media containing selenomethionine as the only source of methionine.

2.3 Protein Purification

2.3.1 Dcc1 containing constructs

Several constructs were amplified from genomic DNA to determine the ones best suited for crystallisation. For the heterotrimer, the subunits Ctf18, Dcc1 and Ctf8 were cloned into the vectors pET-28a, pETDuet and pCDFDuet respectively. An N-terminal His_6 tag and TEV protease site were added to the Ctf18 subunit. For the expression of the Dcc1-Ctf8 sub-complex, Dcc1 and Ctf8 were cloned into pETDuet and pET-28a respectively. In this heterodimer Ctf8 was cloned with the addition of an N-terminal His_6 tag and TEV protease site. For the expression of the C-terminal region of Dcc1 (residue 90-380), Dcc1 was cloned into pET-28a with the addition of an N-terminal His_6 tag and TEV protease site. Constructs were cloned and expressed by the methods presented in section 2.1 and 2.2 respectively.

The Dcc1 constructs were purified by the following procedure. Pellets were re-suspended in Buffer A containing 50 mM Tris pH 7.5, 300 mM NaCl, 10% glycerol, 10 mM imidazole and 0.5 mM TCEP. Cells were then sonicated for 4 minutes using a standard protocol, and centrifuged at 35000 x g, 4 °C for 1 hour. The supernatant was loaded onto a pre-equilibrated 5 mL His-Trap HP (GE Healthcare) column in Buffer A. The column was washed with 100 mL of Buffer A before the protein was eluted in a 75 mL gradient between 10 mM to 0.5 M imidazole. TEV protease was added to the protein to a final concentration of 0.04 $mgmL^{-1}$, 4 °C for 18 hours. To

improve protein purity, the constructs were diluted to an NaCl concentration of 50 mM and loaded onto a pre-equilibrated MonoQ 10/300 GL (GE Healthcare) column in 10 mM Tris pH 7.5, 50 mM NaCl and 0.5 mM TCEP. To elute the protein, a gradient of 100 mL was used between 50 mM to 1 M NaCl. Protein containing fractions were concentrated to 1 mL in a 5000 Da concentrator (Generon) and loaded onto a pre-equilibrated HiLoad 16/60 Superdex 75 column (GE Healthcare) in 10 mM Tris pH 7.5, 150 mM NaCl and 0.5 mM TCEP. Fractions were collected of the protein peak and purity was analysed by SDS-PAGE. Dcc1 constructs were concentrated in a 5000 Da concentrator (Generon) and used for crystallisation trials. Alternatively, for use in biochemical assays the protein was snap frozen in liquid nitrogen in 50 μL aliquots and stored at -80 °C at 10 $mgmL^{-1}$.

2.3.2 Full length Eco1

Full length Eco1 was amplified from genomic DNA and cloned into a pET-28a vector with the addition of an N-terminal His_6 tag and TEV protease site. The cloning and expression of Eco1 was carried out by the methods described in section 2.1 and 2.2. For protein purification frozen Eco1 pellets were re-suspended in Buffer B which contained 50 mM HEPES pH 7.5, 300 mM NaCl, 10 % glycerol, 10 mM imidazole and 0.5 mM TCEP. Cells were sonicated for 4 minutes using a standard protocol and then centrifuged at 35000 x g, 4°C for 1 hour. The supernatant was loaded onto a pre-equilibrated 5 mL His-Trap HP (GE Healthcare) in Buffer B. The column was washed with 100 mL of Buffer B, and Eco1 was eluted in a 75 mL gradient of imidazole between 10 mM to 0.5 M. The cleavage of the His_6 tag was carried out by the addition of TEV protease to a final concentration of 0.04 $mgmL^{-1}$, 4 °C for 18 hours. To further improve protein purity Eco1 was diluted with double distilled water until the NaCl concentration was at 50 mM. This solution was loaded onto a MonoS 10/300 GL (GE Healthcare) column pre-equilibrated in 10 mM HEPES pH 7.5, 50 mM NaCl and 0.5 mM TCEP. Eco1 was eluted in a 100 mL gradient between 50 mM to 1 M NaCl. The protein was collected in 3 mL fractions and concentrated to 1 mL in a 5000 Da concentrator (Generon) at 6000 xg, 4°C. Concentrated Eco1 was loaded onto a pre-equilibrated HiLoad 16/60 Superdex 75 column (GE Healthcare) in 10 mM Tris pH 7.5, 50 mM NaCl and 0.5 mM TCEP. The protein peak was

collected and analysed for purity by SDS-PAGE. Purified Eco1 was concentrated to 10 mgmL^{-1} in a 5000 Da concentrator (Generon) and snap frozen in liquid nitrogen in 50 μL aliquots. Eco1 was stored long term at -80 °C.

2.3.3 xEco2 ACT domain

The xEco2 acetyltransferase (xEco2 ACT) domain (residue 523-702) was amplified from cDNA and cloned into a pET-28a vector with the addition of an N-terminal His_6 tag and TEV protease site. Cloning and protein expression were conducted by the methods presented in Section 2.1 and 2.2 respectively. Frozen cells were re-suspended and sonicated using a standard protocol in 50 mM Tris pH 8.5, 300 mM NaCl, 10 % glycerol, 10 mM imidazole and 0.5 mM TCEP (Buffer C). xEco2 ACT lysate was centrifuged at 35000 x g, 4 °C for 1 hour and the resulting supernatant loaded onto a 5 mL His-Trap HP (GE Healthcare) pre-equilibrated in Buffer C. The column was washed with 100 mL of Buffer C and the acetyltransferase was eluted in a 75 mL gradient of imidazole between 10 mM to 0.5 M. The cleavage of the His_6 tag was carried out by the addition of TEV protease to a final concentration of 0.04 mgmL^{-1}, 4 °C for 18 hours. xEco2 ACT was further purified by diluting the protein until the NaCl concentration was below 50 mM and then loaded onto a pre-equilibrated MonoQ 10/300 GL (GE Healthcare) in 10 mM Tris pH 8.5, 50 mM NaCl and 0.5 mM TCEP. The protein was eluted in a gradient of 100 mL ranging from 50 mM to 1 M NaCl. In the final purification step, the xEco2 ACT was concentrated to 1 mL in a 3000 Da concentrator (Generon) and loaded onto a HiLoad 16/60 Superdex 75 column (GE Healthcare) pre-equilibrated in 10 mM Tris pH 8.5, 150 mM NaCl and 0.5 mM TCEP. Protein containing fractions were analysed for purity by SDS-PAGE. Purified xEco2 ACT was concentrated for use in crystallisation trials in a 3000 Da concentrator (Generon).

2.4 Limited Proteolysis

To optimise protein constructs for crystallisation limited proteolysis was carried out at 4 °C in a buffer of 10 mM Tris pH7.5, 150 mM NaCl and 0.5 mM TCEP. Constructs at 1 mgmL^{-1} were digested by either chymotrypsin or trypsin at a final concentration

of 0.01 mgmL^{-1} and 0.005 mgmL^{-1} respectively. The reaction was mixed and samples were taken after 1 hour, 2 hours and 16 hours. Digested samples were analysed by SDS-PAGE. For further analysis, the gel was transferred onto a nitrocellulose membrane using an iBlot (Novex). The membrane was stained using coomassie blue and protein bands were cut using a sterilised razor. To determine the appropriate cleavage site, N-terminal Edman degradation was employed with intact-mass by MALDI.

2.5 Crystallisation of Protein Constructs

2.5.1 Dcc1 C-terminal domain

The Selenomethionine incorporated C-terminal region of Dcc1 (residue 90-380) was concentrated to 15 mgmL^{-1} and crystallised at 20 °C in 0.2 M potassium sodium tartrate and 18 % PEG 3350. Sitting drops of 2 µL were used, which contained an equal volume of protein and reservoir solution. The crystals were harvested after 2 days in 20 % ethylene glycol and stored in liquid nitrogen. Datasets were collected at two wavelengths on the I03 beamline at Diamond Light Source. To solve the initial structure the AutoSHARP package was used (Bricogne et al., 2003). From these structure factors an initial model was produced using Buccaneer (Cowtan, 2006). To produce the final model, multiple round of refinement and rebuilding were carried out in Coot (Emsley and Cowtan, 2004) and Refmac5 (Murshudov et al., 2011) respectively. Water molecules were added during the final stages of refinement using Coot. Final refinement was carried out against a 2 Å native dataset. Final R-factors of 0.19 for R-work and 0.23 for R-free indicate an accurate model at this resolution. Full statistics for data collection and refinement are shown in Table 2.3, 2.4 and 2.5.

2.5.2 Ctf18$^{666-740}$–Dcc1–Ctf8

The Ctf18$^{666-740}$-Dcc1-Ctf8 complex was concentrated to 35 mgmL^{-1} and crystallised at 20 °C in 4 µL sitting drops with equal volume of reservoir and protein solution. The reservoir contained 0.1 M Bis-Tris propane pH 6.3, 0.2 M NaBr and 17 % PEG 3350. Crystals were harvested after 10 days in 15 % ethylene glycol and stored in liquid

nitrogen. A native dataset was collected on beamline I03 at Diamond Light Source. Molecular replacement was carried out in Phaser (McCoy et al., 2007) using Dcc1$^{90-380}$ as a search model. An initial model was traced using AutoBuild (Adams et al., 2010) and multiple rounds of rebuilding and refinement were carried out in Coot (Emsley and Cowtan, 2004) and Refmac5 (Murshudov et al., 2011) respectively. Coot was used to find water molecules in the last round of refinement. R-factors of 0.20 for R-work and 0.26 for R-free indicate an accurate model at this resolution. Final data collection and refinement statistics are shown in Table 2.6.

Dataset	Peak Dcc1$^{90-380}$
Wavelength	0.9796
Resolution range	79.19 - 2.75 (2.82 - 2.75)
Space group	P 21 21 21
Unit cell	96.69 111.75 112.23 90.00 90.00 90.00
Total reflections	415363 (28991)
Unique reflections	32295 (2349)
Multiplicity	12.9 (12.3)
Completeness (%)	99.9 (99.9)
Mean I/sigma(I)	15.7 (3.3)
Wilson B-factor	55.8
R-merge	0.143 (0.822)
R-meas	0.149 (0.858)
CC1/2	0.996 (0.901)
No.Sites	33
Figure of merit - acentric (centric)	0.35491 (0.11335)
Phasing Power	1.473

Table 2.3. Data Collection Statistics for Dcc1$^{90-380}$ at the Peak Wavelength.

Dataset	Inflection Dcc1$^{90-380}$
Wavelength	0.9797
Resolution range	112.75 - 2.96 (3.04 - 2.96)
Space group	P 21 21 21
Unit cell	96.82 112.02 112.75 90.00 90.00 90.00
Total reflections	336993 (24158)
Unique reflections	26214 (1898)
Multiplicity	12.9 (12.7)
Completeness (%)	99.9 (99.8)
Mean I/sigma(I)	15.8 (3.8)
Wilson B-factor	45.9
R-merge	0.132 (0.746)
R-meas	0.137 (0.778)
CC1/2	0.996 (0.881)
No.Sites	33
Figure of merit - acentric (centric)	0.35491 (0.11335)
Phasing Power	0.391

Table 2.4. Data Collection Statistics for Dcc1$^{90-380}$ at the Inflection Wavelength.

Dataset	Native Dcc1$^{90-380}$
Wavelength	0.92
Resolution range	29 - 2.002 (2.074 - 2.002)
Space group	P 1 21 1
Unit cell	68.682 100.697 83.495 90 100.669 90
Total reflections	288441 (23107)
Unique reflections	71317 (6489)
Multiplicity	4.0 (3.6)
Completeness (%)	0.95 (0.87)
Mean I/sigma(I)	8.43 (1.31)
Wilson B-factor	29.31
R-merge	0.1158 (0.9735)
R-meas	0.1333 (1.138)
CC1/2	0.995 (0.523)
No.Sites	-
Figure of merit - acentric (centric)	-
Phasing Power	-
CC*	0.999 (0.829)
Reflections used in refinement	71312 (6489)
Reflections used for R-free	3560 (324)
R-work	0.1903 (0.2844)
R-free	0.2356 (0.3249)
CC(work)	0.946 (0.641)
CC(free)	0.927 (0.577)
Number of non-hydrogen atoms	8634
macromolecules	8112
Protein residues	1008
RMS(bonds)	0.014
RMS(angles)	1.41
Ramachandran favored (%)	97
Ramachandran allowed (%)	2.5
Ramachandran outliers (%)	0
Rotamer outliers (%)	2.8
Clashscore	5.43
Average B-factor	36.33
macromolecules	36.05
solvent	40.74

Table 2.5. Data Collection and Refinement Statistics for Dcc1$^{90-380}$.

Dataset	Ctf18$^{666-740}$-Dcc1-Ctf8
Wavelength	0.9763
Resolution range	60.64 - 2.29 (2.372 - 2.29)
Space group	P 1 21 1
Unit cell	58.61 164.22 60.64 90 90.55 90
Total reflections	138884 (5577)
Unique reflections	46232 (2926)
Multiplicity	3.0 (1.9)
Completeness (%)	0.90 (0.57)
Mean I/sigma(I)	7.15 (1.31)
Wilson B-factor	43.9
R-merge	0.08522 (0.4577)
R-meas	0.1022 (0.5954)
CC1/2	0.994 (0.732)
No.Sites	-
Figure of merit - acentric (centric)	-
Phasing Power	-
CC*	0.999 (0.919)
Reflections used in refinement	46206 (2922)
Reflections used for R-free	2252 (175)
R-work	0.2018 (0.2967)
R-free	0.2586 (0.3716)
CC(work)	0.950 (0.820)
CC(free)	0.914 (0.591)
Number of non-hydrogen atoms	8784
macromolecules	8518
Protein residues	1050
RMS(bonds)	0.01
RMS(angles)	1.54
Ramachandran favored (%)	96
Ramachandran allowed (%)	2.8
Ramachandran outliers (%)	1.1
Rotamer outliers (%)	2.3
Clashscore	9.77
Average B-factor	52.99
macromolecules	53.07
solvent	50.5

Table 2.6. Data Collection and Refinement Statistics for Ctf18$^{666-740}$-Dcc1-Ctf8.

2.5.3 xEco2 ACT domain

xEco2 ACT domain was crystallised at 20 °C with the protein concentrated to 15 $mgmL^{-1}$. Crystals were obtained of a peptide free form and bound to two different substrate peptides conjugated to CoA, called Ac-CoA-K105 and Ac-CoA-K106 (see Chapter 6). The peptide free form of xEco2 ACT domain crystallised in 0.2 M NaCl, 0.1 M imidazole pH 7 and 0.864 M ammonium phosphate. Crystals were harvested after 3 days in liquid nitrogen with the addition of 30 % ethylene glycol. Ac-CoA-K106 and xEco2 ACT domain crystallised in 40 % ethylene glycol, 0.1 M HEPES pH7.5 and 5 % PEG 3000. Crystals were harvested after 3 days and stored in liquid nitrogen. Ac-CoA-K105 and xEco2 ACT domain crystallised in 0.18 M $MgCl_2$, 10% glycerol, 27% 2-propanol and 0.09 M HEPES pH 7.5. These crystals were harvested after 5 days in 25 % ethylene glycol and stored in liquid nitrogen. Data was collected on beamline I02 at Diamond Light Source for Ac-CoA-K106 and the peptide free form of xEco2 ACT domain. For Ac-CoA-K105 data was collected on beamline I03 at Diamond Light Source.

The structure of Ac-CoA-K106 was solved by molecular replacement conducted in Phaser (McCoy et al., 2007) using hEsco2 (Kouznetsova et al., 2016) as the search model. AutoBuild (Adams et al., 2010) was then used to trace a start model and multiple rounds of rebuilding and refinement were conducted in Coot (Emsley and Cowtan, 2004) and Phenix.Refine (Adams et al., 2010) respectively. Full data collection and refinement statistics are shown in Table 2.7. Final R-factors of 0.21 for R-work and 0.24 for R-free indicate an accurate model at this resolution. The Ac-CoA-K105 structure was solved by molecular replacement in Phaser (McCoy et al., 2007) using xEco2 ACT domain as a search model. Multiple rounds of rebuilding and refinement were carried out in Coot (Emsley and Cowtan, 2004) and Phenix.Refine (Adams et al., 2010) respectively. Full data collection and refinement statistics are shown in Table 2.8. Final R-factors of 0.21 for R-work and 0.23 for R-free indicate an accurate model at this resolution. In both cases the peptide was built in the final stage of refinement after which water was added using coot. The peptide free structure of xEco2 ACT domain was solved by molecular replacement in Phaser (McCoy et al., 2007) using xEco2 ACT domain as a search model. Multiple rounds of refinement and rebuilding were carried out in Phenix.Refine (Adams et al., 2010)

and Coot (Emsley and Cowtan, 2004) respectively. Full data collection and refinement statistics are shown in Table 2.9. Final R-factors of 0.22 for R-work and 0.27 for R-free indicate an accurate model at this resolution.

Dataset	Ac-CoA-K106
Wavelength	0.9795
Resolution range	59.16 - 1.99 (2.061 - 1.99)
Space group	P 1 21 1
Unit cell	107.236 57.7683 70.9545 90 89.9746 90
Total reflections	197266 (19580)
Unique reflections	59859 (5936)
Multiplicity	3.3 (3.3)
Completeness (%)	0.98 (1.00)
Mean I/sigma(I)	8.25 (1.65)
Wilson B-factor	31.92
R-merge	0.06494 (0.7646)
R-meas	0.07758 (0.9127)
CC1/2	0.998 (0.615)
CC*	0.999 (0.873)
Reflections used in refinement	58984 (5707)
Reflections used for R-free	2817 (270)
R-work	0.2080 (0.3658)
R-free	0.2374 (0.3866)
CC(work)	0.955 (0.811)
CC(free)	0.939 (0.721)
Number of non-hydrogen atoms	6005
Macromolecules	5497
Ligands	244
Protein residues	689
RMS(bonds)	0.007
RMS(angles)	0.77
Ramachandran favored (%)	97
Ramachandran allowed (%)	2.7
Ramachandran outliers (%)	0.6
Rotamer outliers (%)	3.8
Clashscore	3.67
Average B-factor	45.96
Macromolecules	46.41
Ligands	35.36
Solvent	46.39
Number of TLS groups	23

Table 2.7. Data Collection and Refinement Statistics for xEco2-K106.

Dataset	Ac-CoA-K105
Wavelength	0.9795
Resolution range	46.97 - 2.3 (2.382 - 2.3)
Space group	P 1 21 1
Unit cell	57.3015 60.6099 66.9053 90 99.6344 90
Total reflections	131452 (13235)
Unique reflections	20265 (2010)
Multiplicity	6.5 (6.6)
Completeness (%)	1.00 (1.00)
Mean I/sigma(I)	6.78 (2.02)
Wilson B-factor	31.9
R-merge	0.2701 (1.834)
R-meas	0.2943 (1.992)
CC1/2	0.98 (0.639)
CC*	0.995 (0.883)
Reflections used in refinement	20227 (2001)
Reflections used for R-free	988 (99)
R-work	0.2102 (0.3180)
R-free	0.2297 (0.3884)
CC(work)	0.948 (0.841)
CC(free)	0.909 (0.684)
Number of non-hydrogen atoms	2999
Macromolecules	2719
Ligands	124
Protein residues	348
RMS(bonds)	0.005
RMS(angles)	0.63
Ramachandran favored (%)	95
Ramachandran allowed (%)	4.5
Ramachandran outliers (%)	0.3
Rotamer outliers (%)	2.4
Clashscore	3.71
Average B-factor	44.17
Macromolecules	44.67
Ligands	33.44
Solvent	44.09

Table 2.8. Data Collection and Refinement Statistics for xEco2-K105.

Dataset	Peptide free
Wavelength	0.9795
Resolution range	28.83 - 2.978 (3.084 - 2.978)
Space group	P 21 21 21
Unit cell	58.475 66.287 109.763 90 90 90
Total reflections	55624 (4904)
Unique reflections	9148 (860)
Multiplicity	6.1 (5.7)
Completeness (%)	0.99 (0.96)
Mean I/sigma(I)	7.70 (1.52)
Wilson B-factor	83.52
R-merge	0.2968 (1.159)
R-meas	0.3252 (1.27)
CC1/2	0.968 (0.449)
CC*	0.992 (0.787)
Reflections used in refinement	9089 (837)
Reflections used for R-free	908 (83)
R-work	0.2247 (0.3684)
R-free	0.2669 (0.3897)
CC(work)	0.956 (0.547)
CC(free)	0.899 (0.566)
Number of non-hydrogen atoms	2725
Macromolecules	2629
Ligands	96
Protein residues	340
RMS(bonds)	0.007
RMS(angles)	0.88
Ramachandran favored (%)	94
Ramachandran allowed (%)	6.1
Ramachandran outliers (%)	0.3
Rotamer outliers (%)	5.7
Clashscore	7.06
Average B-factor	72.81
Macromolecules	73.25
Ligands	60.61
Solvent	-

Table 2.9. Data Collection and Refinement Statistics for xEco2-Peptide Free.

2.6 EMSA

2.6.1 Polyacrylamide EMSA

To determine the binding of Dcc1 to DNA an electrophoretic mobility shift assay (EMSA) was performed by native-PAGE with both a 25-mer oligo d(T) ssDNA and 21-mer dsDNA, which had a sequence of TCTCCACAGGAAACGGAGGGGT. Both substrates were commercially synthesised (SIGMA) to contain a 6-FAM fluorescent tag at the 5' end of the sequence. To produce dsDNA an oligo of complementary sequence was ordered without a fluorescent tag. Both oligos were then mixed in an equal molar ratio and heated to 95 °C. The DNA was then annealed by controlled cooling over 70 minutes to 25 °C. The dsDNA was purified to remove any ssDNA by loading the sample onto a pre-equilibrated Resource 15Q column (GE Healthcare) in 50 mM phosphate buffer pH 7.5. The dsDNA was eluted in a 120 mL gradient between 0 to 1 M KCl.

A 6 % non-denaturing polyacrylamide gel made from bis-acrylamide (Fisher) with a 19:1 ratio of bis-acrylamide to crosslinker in 1 x Tris/borate/EDTA (TBE) was pre-run at 3 mA for 60 minutes at 4 °C in 0.5 x TBE. 1 μM DNA was used for the binding reactions and the protein was titrated at increasing concentrations in 10 mM Tris pH 7.5, 150 mM NaCl and 0.5 mM TCEP. The samples were loaded onto the gel and run for 75 minutes at 5 mA, 4 °C. The gel was visualised on a Typhoon 9400 excited at 488 nm at 500 V and the emission was detected at 520 nm.

2.6.2 Agarose EMSA

To determine the binding affinity of the head domain of cohesin to DNA a human head complex was expressed and purified from insect cells using standard protocols ($Smc1^{\Delta 226\text{-}1003}$, $Smc3^{\Delta 230\text{-}964}$, full-length Scc1 and Scc3). Mutant Smc3 was made using the method described in section 2.1. A 700 base pair dsDNA was amplified from a plasmid and used at a final concentration of 14 nM. The protein was titrated at increasing concentration in 50 mM Tris pH 8.0, 50 mM NaCl and 0.5 mM DTT. The samples were loaded onto a 0.8 % agarose gel, run for 20 minutes at 80 V in TBE. UV light was used to visualize an ethidium bromide gel.

2.7 Glycerol Gradient

To analyse the interaction between Pol2 and the $Ctf18^{C}$-Dcc1-Ctf8 heterodimer a 20-50 % glycerol gradient was prepared in 20 mM HEPES pH 7.5, 25 mM NaCl and 1 mM TCEP; using a gradient master (Biocomp) in a 13 x 51 mm polyallomer centrifuge tubes (Beckman). The prepared gradient was left to cool to 4 °C for 1 hour before a 120 µL solution containing a final concentration of proteins at 9 µM in 20 mM HEPES pH 7.5, 25 mM NaCl and 1mM TCEP was added to the top of the glycerol gradient. Ultracentrifugation was conducted on the samples at 44000 rpm, 4 °C for 18 hours. After which, the gradient was separated into 400 µL fractions and analysed by SDS-PAGE.

2.7.1 Expression and purification of Pol2

The N-terminal region of Pol2 (residue 1-544) was cloned from *S. cerevisiae* genomic DNA into pET28a with the addition of an His_6 N-terminal tag and TEV protease site using the procedure presented in Section 2.1. The expression was carried out by the same method presented in Section 2.2.

Pol2 pellets were re-suspended in Buffer A. Cells were then sonicated for 4 minutes using a standard protocol and the sample was centrifuged at 35000 x g, 4 °C for 1 hour. The supernatant was loaded onto a pre-equilibrated 5 mL His-Trap HP (GE Healthcare) column in Buffer A. The column was washed with 100 mL of Buffer A before the protein was eluted in a 75 mL gradient from 10 mM to 0.5 M imidazole. TEV protease was added at to a final concentration of 0.04 $mgmL^{-1}$, 4 °C for 18 hours. To improve protein purity Pol2 was diluted to an NaCl concentration of 50 mM and loaded onto a pre-equilibrated MonoQ 10/300 GL (GE Healthcare) column in 10 mM Tris pH 7.5, 50 mM NaCl and 0.5 mM TCEP. To elute pol2, a gradient of 100 mL was used of NaCl between 50 mM to 1 M. Protein containing fractions were concentrated to 1 mL in a 20000 Da concentrator (Generon) and loaded on a pre-equilibrated HiLoad 16/60 Superdex 75 column (GE Healthcare) in 10 mM Tris pH 7.5, 150 mM NaCl and 0.5 mM TCEP. Fractions were collected and purity was analysed by SDS-PAGE. Pol2 containing fractions were concentrated to 10 $mgmL^{-1}$and snap frozen in liquid nitrogen in 50 µL aliquots, which were stored at -80 °C.

2.8 Consensus Motif

2.8.1 Acetylation of cohesin subunits *in vitro*

Acetylation of cohesin subunits by Eco1 was carried out with 70 µM and 0.7 µM of Eco1 and cohesin subunits (Smc1-Smc3-Scc1-Scc3, Scc2-Scc4, Pds5 and Wapl) respectively in 50 mM Tris pH 8, 50 mM NaCl, 0.5 mM NaCl, 0.5 mM DTT, 5 mM $MgCl_2$, 1.5 mM Acetyl-CoA and 1 mM ATP. The acetylation reaction was heated to 37 °C. After 90 minutes the reactions were stopped by the addition of SDS-loading buffer. Samples were analysed by SDS-PAGE and the proteins were visualised by coomassie blue staining. Appropriate bands were cut with a sterilized razor, and samples analysed for post-translational modification by MS. In addition, a time course experiment was carried out with the core cohesin complex in the same conditions as above. However, additional samples were taken at 0, 10, 20, 30 and 40 minutes.

2.8.2 Analysis of peptides and generation of a consensus motif

After background acetylation by insect cells had been taken into consideration a consensus acetylation motif was generated using the iceLogo server (Colaert., 2009). Peptides acetylated by Eco1 were manually split into three categories of forward, reverse or no direction (see Chapter 6 for further details).

Chapter 3. Theory of Crystallography

The technique called x-ray crystallography helps improves our understanding of proteins and their function by determining their atomic structure. The first step to solving a protein structure is to obtain a pure protein at relatively large quantities (~5 mg of protein). The protein then needs to be stable to enable the formation of protein crystals. This is usually the limiting step in x-ray crystallography (Saccardo et al., 2016).

Once a protein crystal is grown the crystallisation condition needs to be optimised to improve the crystal's diffraction. This can be a challenging process, usually, taking many rounds of refinement. The best crystals are harvested in liquid nitrogen, and then shot with high powered x-rays. Once a full dataset is acquired the phases need to be determined to allow an initial electron density map to be calculated. There are three major methods used to solve the phase problem known as isomorphous replacement, anomalous diffraction and molecular replacement (Rhodes, 2010). The growing number of atomic co-ordinates deposited in the protein data bank (PDB) has led to molecular replacement being the most common method used today to determine phase information.

Once an initial electron density map has been calculated the map is then refined using standard procedures before a polypeptide chain is built. Multiple rounds of rebuilding and refinement are carried out to produce the final model. To analyse the validity of the solved structure statistical tools are employed throughout the refinement.

Once the final model has been produced the atomic structure is analysed to help determine protein function. Multiple databases and online servers exist to do this. For instance, the ConSurf server (Ashkenazy et al., 2016), which analyses the surface conservation of a protein; or the Dali server (Holm and Rosenström, 2010), which compares a new structure against co-ordinates in the PDB.

In this chapter the practical methods of producing crystal are discussed. Then the theory of x-ray crystallography is described in detail starting with the symmetry of

crystals. Next, the process of turning the diffraction images into an electron density map is presented; along with methods to solve the phase problem. Finally, the process of refinement is examined.

3.1 Protein Crystallisation

3.1.1 Properties of crystals

Protein crystals are defined as solid structure which are made up of a basic repeating unit (Rhodes, 2010), and they are different from amorphous precipitation as they are ordered. Crystal which diffract to a high resolution have well-ordered identical unit cells. However, protein crystals are usually made from an imperfect arrangement, which is measured by the mosaic spread (Rhodes, 2010). The mosaic spread causes the diffracted x-ray beam to be a cone shaped rather than a regular beam. Therefore, diffraction needs to be recorded over a small angle to measure all reflections.

Protein crystals can be extremely fragile as unlike salt crystals they are held together through hydrogen bonds between the protein and water molecules. This also means the diffraction has to be recorded in the mother liquor because dehydrating the protein crystal will result in the breaking of the crystal lattice.

3.1.2 Growing protein crystals

The formation of protein crystal only occurs when the molecule is stable and can be purified to a high standard. Once the protein is pure it is mixed with a precipitant at a concentration just below the concentration required to precipitate the protein (Figure 3.1) (Khurshid et al., 2014). The concentration of both protein and precipitant is then gradual increased by the removal of water, which causes the protein to precipitate. This process hopefully leads to the formation of protein crystals (García-Ruiz, 2003). The slower protein precipitation occurs the more likely the molecule will crystallise rather than form amorphous precipitation. Standard protein precipitants include molecules like polyethylene glycol, ammonium sulphate, ethanol or MPD. There are two stages to crystal growth called nucleation and growth. Firstly, nucleation occurs, which is the establishment of a molecular clusters; and secondly,

growth of the crystal follows, which is the addition of proteins to this initial micro-crystal.

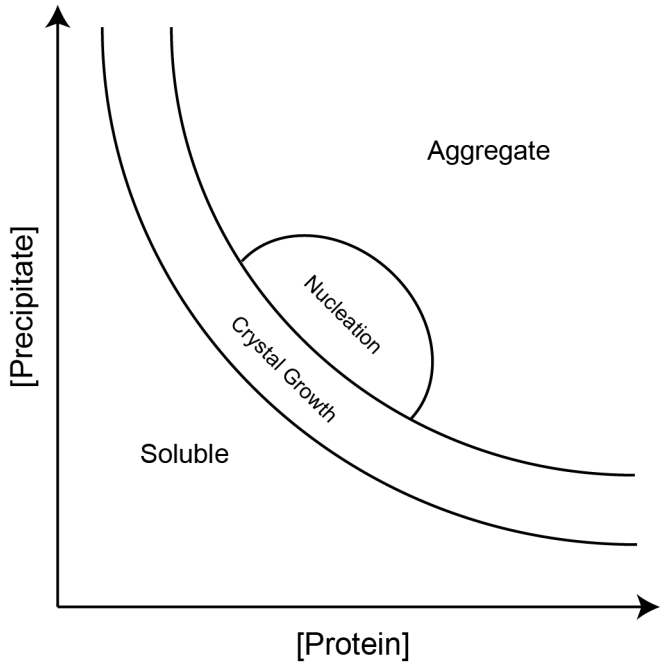


Figure 3.1. Phase Diagram of Protein Against Precipitant.

Normally it is possible to crystallise the protein in different crystal forms, which could vary in their reproducibility, diffraction, size, shape or symmetry. The different crystal forms are then optimised to produce crystals which can diffract to a high-resolution. Sometimes when a model for molecular replacement is not available the phases needs to be determined experimentally. One method to do this requires the production of multiple heavy metal derivatives. These can be made by co-crystallisation or soaking the crystal in a heavy metal solution.

The optimisation of crystals occurs through varying a variety of conditions; for example, the temperature, precipitant concentration, pH, protein concentration, drop size or salt concentration (McPherson and Cudney, 2014). Varying the above conditions can sometimes not improve the overall diffraction quality and other procedures can be attempted such as crystal seeding; the use of small molecules; detergents or antibodies (McPherson and Cudney, 2014). Finally, if crystals still cannot be improved the protein construct may need to be re-designed.

3.1.3 Crystal data-collection

Once high quality crystals have been grown they need to be harvested. Old methods include room temperature shooting, which involved transferring the crystal along with the mother liquor to a glass capillary. However, cryo methods to harvest and shoot the protein crystals are now normally used (Pflugrath, 2015). To do this, the crystal firstly needs to be cryo-protected to stop the formation of ice crystals by the addition of glycerol, ethylene glycol or other similar small molecules. Once the crystal is cryo-protected it is then shot at liquid nitrogen temperatures (-196 °C). There are many advantages of cryo data-collection; for instance, at lower temperatures the crystal lattice is more stable; has greater resistance to radiation damage; and crystals can easily be transported in dry-shippers (Pflugrath, 2015).

3.2 Crystal Symmetry

3.2.1 Unit cell symmetry

All crystals can be defined by their unit cell, which is the smallest single repeating unit of a crystal. Any unit cell is described by having three axes called a, b and c; and three angles α, β and γ. There are a number of different unit cells defined by their angles and relative axes lengths. For instance, when $a \neq b \neq c$ and $\alpha \neq \beta \neq \gamma$ the unit cell is called a triclinic cell. A monoclinic cell is defined when $a \neq b \neq c$ and $\alpha = \gamma$, $\beta > 90$ °. A hexagonal cell has the criteria that $a = b$, $\alpha = \beta$ and $\gamma = 120$ °. For a tetragonal or cubic cell all three angles equal 90 °. However, for a cubic cell all axes are equal in a length ($a = b = c$) whilst for a tetragonal cell the c axis is different ($a = b \neq c$) (Rhodes, 2010).

3.2.2 Bragg's law

The angle of diffraction from a crystal lattice is related to sets parallel planes (Bragg, 1913). This model describes a reflection being caused by a set of parallel plane in the crystal. Each set of parallel planes in a crystal are described by three indices (miller indices) h, k and l; separated with an inter-planer spacing d_{hkl}. The h, k and l define the number of planes in the x-, y- and z-direction respectively between lattice

point. For instance, if a set of parallel plane has the indices (300) then there would be 3 planes between lattice points in the x-direction.

Bragg's law states that a reflection can only be observed from a set of parallel planes if the x-ray travels an integral distance of the wavelength (λ) between the planes (Equation 1). Only when this is satisfied do reflected x-rays construct in phase leading to a strong reflection. However, if the distance travelled is not an integral number of wavelengths then the reflected x-rays will be out of phase and add up destructively.

$$n\lambda = 2d_{hkl}\sin\theta$$

(Eq. 1)

Equation 1 (Bragg's law) can be derived from drawing a set of parallel planes between lattice points (Figure 3.2). For instance, if two parallel x-rays of wavelength λ strike the parallel planes at two different lattice points the x-rays will reflect at the same angle of incidence (θ). From the diagram in Figure 3.2:

$$\sin\theta = \frac{BC}{AB}$$

(Eq. 2)

Therefore:

$$\mathrm{AB}\sin\theta = BC$$

(Eq. 3)

As BC = d_{hkl} and the extra distance travelled by R_2 is 2BC then:

$$2\mathrm{BC} = 2\mathrm{d_{hkl}}\sin\theta$$

(Eq. 4)

For a reflection to be observed 2BC has to equal an integral number of λ, which derives Equation 1.

Bragg's law means that the number of reflection depend on the parallel planes in a unit cell. Therefore, the position and number of spots in a diffraction pattern is related to the unit cell dimensions and not the cell contents.

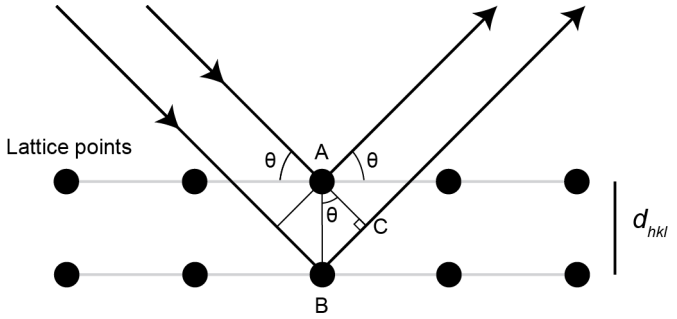


Figure 3.2. Bragg's law Derived From Diffraction From Parallel Planes.

3.2.3 The reciprocal lattice

The reflections produced from a set of parallel planes are present in reciprocal space. Reciprocal space is inversely proportion to real space so that the volume of the reciprocal unit cell ($V_{reciprocal}$) relates to volume of real unit cell (V_{real}) by Equation 5 (Rhodes, 2010).

$$V_{reciprocal} = \frac{1}{V_{real}}$$

(Eq. 5)

Therefore, if the real unit cell edges are in Å the reciprocal cell edges would have the units $Å^{-1}$. It also means a small real unit cell will have a large reciprocal lattice and a large real unit cell would have a small reciprocal lattice. For protein crystals, as the unit cell dimension are in Å they have an inherently large reciprocal lattice.

3.2.4 Space groups

Crystal unit cells can be further broken down into symmetrically related units called the asymmetric unit. For instance, in a protein crystal in any one-unit cell there are normally multiple proteins which can be related by symmetry operations like a translation or rotation. There are more complex symmetry operations such as a screw axis, which are formed from a rotation and a translation. The combination of symmetry elements in a unit cell gives rise to its space group. There are 230 possible space group; however, the chiral nature of proteins limits the use of the reflection

symmetry operator, which leads to protein crystals only having 65 possible space groups (Hahn, 2005).

Space groups are defined by a letter and a number such as, $P2_1$. The letter indicates the lattice type or Bravais lattice, and the numbers reveal the symmetry operation. There are thirteen Bravais lattice types, which are determined first. In a cubic system there are three lattices known as Primitive (P), Internal (I) and Face-centred (F) (Figure 3.3). Usually for a given crystal there are a number of lattice types to choose from, and the standard procedure is to choose the highest symmetry Bravais lattice. For instance, a F lattice has higher symmetry than a P lattice. In the $P2_1$ spacegroup, the 2_1 indicates that the protein has a two-fold screw axis. This means the molecule rotates 180 ° and translates half the axial length. Alternatively, a 3_1 would mean there is a three-fold screw axis and the molecule would rotate 120 ° and move a third of the axial length (Hahn, 2005).

The space group can be determined by analysing the reflections intensities from sets of parallel planes. The first step is to calculate the lattice type and unit cell dimensions. It is then possible to determine the precise space group by defining the additional symmetry operators. In some cases, this can be revealed from systematic absences. For instance, a two-fold screw axis along the b-axis will cause missing reflections from the hkl planes when k is an odd number (Hahn, 2005).

The variability of crystals in both their size and strength of diffraction means that after reflections have been collected the frames need to be scaled. The collection of the same reflection in multiple frames can aid this scaling process.

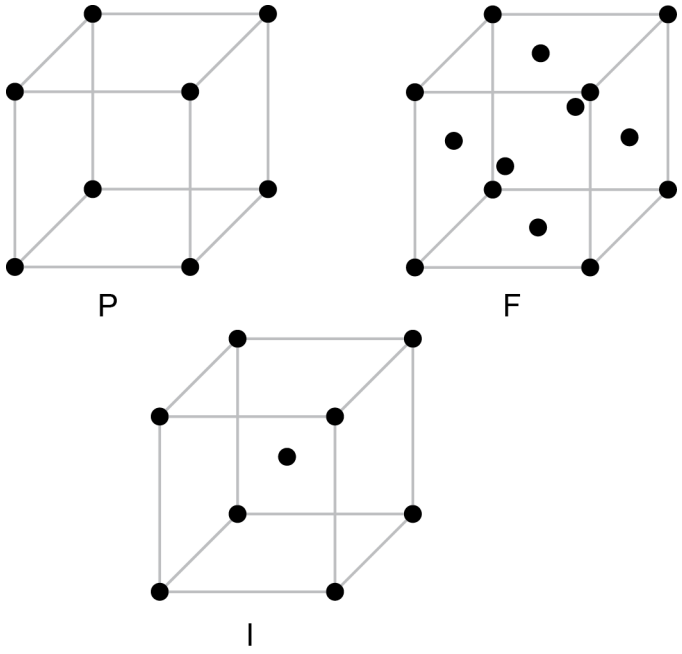


Figure 3.3. Cubic Lattices.

The dots signify lattice points, and the letters P, I and F represent the Primitive, Internal and Face-centred cubic lattice types respectively.

3.3 Calculating the Electron Density

The electron density of a unit cell is calculated from the reflection intensities recorded from a diffraction pattern and the phases, which cannot be obtained from the reflections. The electron density repeats itself; therefore, it can be approximated by a Fourier sum, and represented as a periodic function. To understand the structure factor (F_{hkl}) and the resulting electron density equation $\rho(x, y, z)$ the theory of waves will firstly be described.

3.3.1 Mathematical description of waves

The trigonometry functions of sin or cos can be used to describe any periodic function in one-dimension (Equation 6 and 7). Where F, h and α represent the amplitude, frequency and phase respectively.

$$f(x) = \mathrm{F}\cos 2\pi(hx + \alpha)$$

(Eq. 6)

Or:

$$f(x) = \mathrm{F}\sin 2\pi(hx + \alpha)$$

(Eq. 7)

A complex wave can be approximated by a Fourier sum (Rhodes, 2010). Therefore, no matter how complex the wave is by selecting the correct value for the frequency, amplitude and phase it is possible to describe any wave (Rhodes, 2010).

$$f(x) = \sum_{h=0}^{n} \mathrm{F_h}\cos 2\pi(hx + \alpha_h)$$

(Eq. 8)

Furthermore, waves can be described by a complex number (Equation 9). The phase is implicit in this complex number equation.

$$f(x) = \sum_{h=0}^{n} \mathrm{F_h}[\cos 2\pi(hx) + i \sin 2\pi(hx)]$$

(Eq. 9)

Equation 9 can further be simplified as:

$$e^{i\theta} = i \sin\theta + \cos\theta$$

(Eq. 10)

Therefore:

$$f(x) = \sum_{h} F_h e^{-2\pi i(hx)}$$

(Eq. 11)

Equation 11 represents a complex periodic function in one direction. Furthermore, the Fourier sum can be used to describe a wave in three dimensions (Equation 12).

$$f(x,y,z) = \sum_{h}\sum_{k}\sum_{l} F_{hkl} e^{-2\pi i(hx+ky+lz)}$$

(Eq. 12)

Equation 12 can be used to represent the electron density in three dimensions as will be shown in the next sections.

3.3.2 Fourier transform

Another mathematical function is called a Fourier transform (FT) (Rhodes, 2010). This is useful as it allows the movement between the real and reciprocal space. The principle behind the FT is calculating the reciprocal function of a complex sum. In this way, the FT is reversible and carrying it out twice on any function will result in the original function being obtained.

3.3.3 Description of the structure factor

A reflection can be represented by a Fourier sum, which is called the structure facture (F_{hkl}). The contribution of a single atom to the structure factor can be described when taking into account the atoms scattering factor (f_j) (Rhodes, 2010).

$$f_{hkl} = f_j e^{2\pi i(hx_j+kx_j+l_j)}$$

(Eq. 13)

However, any single structure factor is the sum of contribution from every atom in the unit cell, and is described by a Fourier sum.

$$F_{hkl} = \sum_{j=1}^{n} f_j e^{2\pi i(hx_j+kx_j+l_j)}$$

(Eq. 14)

The term f_j is equivalent to the average electron density. Therefore, the structure factor can also be written in terms of $\rho(x, y, z)$. Although $\rho(x, y, z)$ is an average density over a set volume by using an infinitely small volume it is possible to obtain the precise value of $\rho(x, y, z)$ at any given co-ordinate. This can be done by integrating all values of x, y and z over the unit cell (v) in the Fourier sum, which gives the following equation.

$$F_{hkl} = \int_V \rho(x,y,z) e^{2\pi i(hx+kx+lx)}\, dv$$

(Eq. 15)

Carrying out a FT on Equation 15 derives ρ(x, y, z) from the structure factors.

$$\rho(x,y,z) = \frac{1}{V}\sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i(hx+ky+lz)}$$

(Eq. 16)

The electron density can be determined from the periodic function presented in Equation 16. To do this, the structure factors need to be calculated, which can be obtained from intensity of the reflection as F_{hkl} is proportional to the square root of the measured reflection intensities (I_{obs}) (Rhodes, 2010). In addition, the frequency and phase of each term is required. The frequency can be can be calculated from the indexed reflections as the frequency is inversely proportional to the inter-planner spacing of the plane that caused the reflection. However, the phase cannot be measured from the diffraction pattern and further information is required to calculate initial phase estimates.

Once initial phases are determined multiple rounds of phase improvement are carried out to enhance the overall electron density. This is done, firstly, by calculating ρ(x, y, z) based on initial phase estimates α_{calc} and the reflection intensities (I_{obs}). The phases are then improved by identifying the position of atoms or using techniques such as solvent flattening, which helps identify solvent boundaries. Structure factors are then re-calculated to determine improved phases. This process is repeated until I_{calc} reaches I_{obs} (Rhodes, 2010).

3.4 Solving the Phase Problem

To obtain ρ(x, y, z) from equation 16, the phase of each structure factor in the Fourier series needs to be estimated. The electron density can then be represented as a contour map. Initial phases are usually estimates, which are improved through multiple rounds of refinement.

In this section the mathematical relationship between ρ(x, y, z) and the phase of each reflection will be described. Then the three common methods of obtaining phases will be discussed.

3.4.1 Representation of ρ(x, y, z) with phases

In Equation 16. ρ(x, y, z) is represented in terms of structure factors (F_{hkl}) which have a phase, intensity and frequency. Both the intensity and frequency of a reflection can be measured from the diffraction pattern. To reveal how phase information is obtained it is useful to represent ρ(x, y, z) in another form (see below).

Structure factors are described as being complex vectors with a real component known as the intensity and an imaginary comment termed the phase (Figure 3.4). In this way the phase represents an angle between the vector and the real axis. A useful representation of phases can be derived from the diagram as follows. Where |F| is proportional to the square root of the amplitude.

$$\sin 2\pi\alpha = \frac{|A|}{|F|}$$

(Eq. 17)

and:

$$\cos 2\pi\alpha = \frac{|B|}{|F|}$$

(Eq. 18)

Representing Equation 17 and 18 as a complex number gives rise to the following equation for a structure factor:

$$F = |F|(\cos 2\pi\alpha + i \sin 2\pi\alpha)$$

(Eq. 19)

Therefore:

$$F = |F|e^{2\pi i\alpha}$$

(Eq. 20)

Substituting F into Equation 16 gives rise to the following:

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}|F_{hkl}|e^{2\pi i\alpha_{hkl}}e^{-2\pi i(hx+ky+lz)}$$

(Eq. 21)

Or:

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}|F_{hkl}|e^{-2\pi i(hx+ky+lz-\alpha_{hkl})}$$

(Eq. 22)

Equation 22 describes ρ(x, y, z) in terms of amplitude, phase and frequency (Rhodes, 2010). Using this equation, it is possible to determine an initial electron density map. The methods used to obtain phase information shall be discussed below.

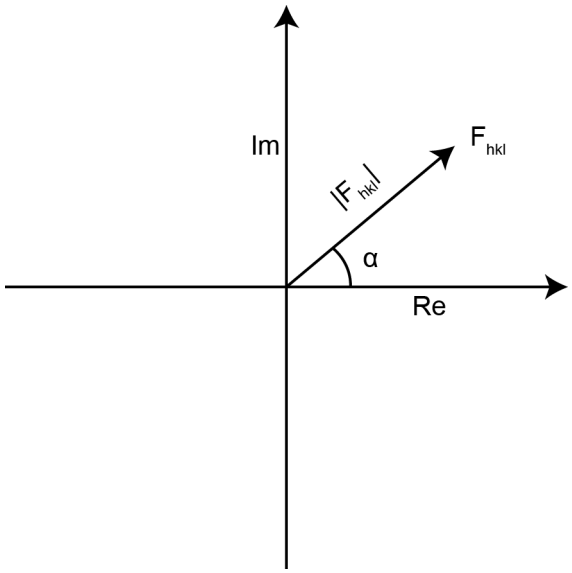


Figure 3.4. Structure Factors Represented as Complex Numbers.

3.4.2 Isomorphous replacement

One method of experimental phase determination is called isomorphous replacement or the heavy atom method (Taylor, 2010). The principle behind this technique is to incorporate atoms into the unit cell and observe the effect on the diffraction pattern. The reflections reflect the lattice; therefore, adding atoms will cause a change in the observed reflections. For the change to be observable the

incorporated atom has to diffract x-rays more strongly compared to the atoms in a protein. Atoms with a high molecular mass are, therefore, chosen as good candidates for isomorphous result.

To obtain crystals with heavy atoms incorporated into the unit cell, crystals can either by soaked or co-crystallised with a heavy atom of choice. Common ions include platinum, gold, mercury or uranium. Heavy metals are usually poisonous to the crystal lattice; as a result, the concentration and type of ion used usually requires extensive testing (Sun et al., 2002). Furthermore, the heavy metal has to bind to the protein in the same position in every unit cell at a high occupancy otherwise the contribution to the diffraction pattern from the heavy metal will be unobservable. The heavy atom also cannot alter the unit cell dimensions or crystal symmetry else the difference in the diffraction pattern will not solely be due to the addition of a heavy atom to the unit cell.

Once two data-sets of sufficient quality are obtained, the next step in gaining phase information is to find the location of the heavy atoms. To do this, the difference between the intensity of each reflections is calculated:

$$|F_h|^2 = (\left|F_{hp}\right| - |F|_p)^2$$

(Eq. 23)

Where $|F_h|$, $|F_{hp}|$ and $|F_p|$ is the intensity of reflections caused by the heavy atom, heavy atom and protein, and protein alone respectively. Equation 23 is squared to make the observation positive.

The next step is to calculate a Patterson function, which is another Fourier sum (Taylor, 2010). A Patterson function is a different way to represent the unit cell content. However, in this description the phase of each structure factor is not required. A basic Patterson function is described as follows:

$$P(u,v,w) = \frac{1}{V}\sum_h \sum_k \sum_l |F_{hkl}|^2 e^{-2\pi i(hu+kv+lw)}$$

(Eq. 24)

In Equation 24, *P (u, v, w)* is a mathematical representation of the vectors between atoms in a unit cell. For a protein molecule the resulting vector map is too complicated to interpret. However, for a unit cell which contains a few atoms their position can readily be identified. The location of peaks in specific locations in the Patterson map can help identify symmetry elements, such as a rotation axis, which makes finding the position of atoms easier. For proteins soaked in heavy atoms computer programmes usually trial several solutions to determine the correct location of the heavy atoms.

Combining Equation 24 with 23 gives rise to a Patterson function of just the heavy atoms ($|F_h| = \Delta F$).

$$\Delta P(u,v,w) = \frac{1}{V}\sum_h \sum_k \sum_l \Delta {F_{hkl}}^2 e^{-2\pi i(hu+kv+lw)}$$

(Eq. 25)

It is therefore possible to find the location of the heavy atom in the unit cell and their resulting phases from two datasets.

One of the potential issues with the use of a Patterson function is that it is unable to determine the difference between inverted images, which is known as the hand problem. As a result, the Patterson function gives rise to two solutions called the left and right hand. If high resolution data is available both hands are taken forward and the electron density map is refined, which should give one solution with a significantly better electron density.

Once the phases of the heavy atoms have been solved an estimate of the phase of the protein needs to be determined. This can be done with the aid of a Harker diagram (Figure 3.5) in combination with the following equation:

$$F_p = F_{ph} - F_h$$

(Eq. 26)

Using a Harker diagram the vector representing the structure factor $-F_h$ is plotted. Next, a circle is drawn with a radius equal to $|F_{ph}|$, with its centre at the end of the vector representing $-F_h$. Finally, a circle with a radius equal to $|F_p|$ is plotted at the

origin (Figure 3.5). The points these two circles intersect gives rise to the phase of F_p. It is common for the circles to intersect more than once meaning an exact value for the phase cannot be determined. If the intersections lie close together then an average is usually taken, which is normally enough for an initial phase estimate. However, sometimes this is not feasible and it may be necessary to determine phase information by the use of an additional derivatives.

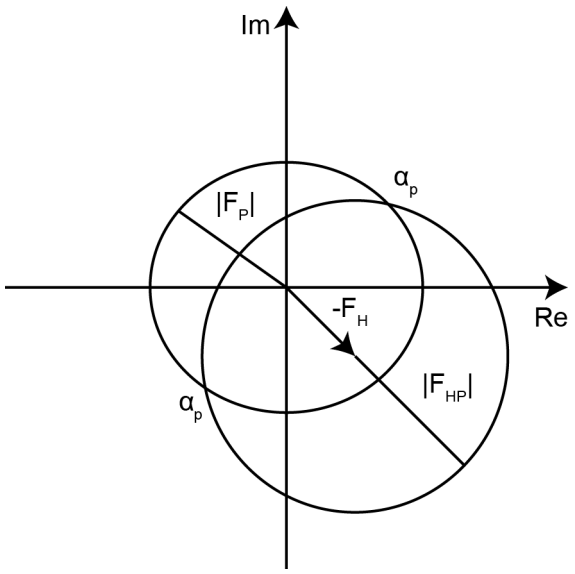


Figure 3.5. Solving Phases Using a Harker Diagram.

3.4.3 Anomalous scattering

Another method to determine phase information is to use the fact that atoms absorb photons at a specific wavelength and then re-admit them with a different phase (Taylor, 2010). Atoms absorb x-rays strongly at a wavelength just lower than their emission wavelength, which is called the absorption edge. The wavelength of x-rays is around the absorption edge of heavy atoms. Therefore, again a second derivative is usually required to obtain an anomalous signal unless the protein already contains an appropriate atom such as in zinc finger domains.

The development of producing proteins with selenomethionine incorporated into the polypeptide chain has made anomalous scattering one of the most common experimental phasing methods. For instance, if the crystallised protein was recombinantly expressed in *Escherichia coli* then it is possible to grow the cells in

minimal media where selenomethionine is the only source of the amino acid methionine. In this way, selenomethionine is incorporated into the protein. The selenium atom can then act as an anomalous diffractor if the correct wavelength is selected (Walden, 2010).

Once a suitable crystal derivative has been obtained a dataset needs to be collected at the heavy atoms absorption edge. The anomalous signals can be identified from observing Friedel pairs. Friedel law states that the reflections caused by planes hkl and -hkl are equal. However, Friedel law is no longer true when an anomalous diffractor is present (Taylor, 2010). It is possible to generating a Patterson function from the difference in Friedel pairs using Equation 27.

$$F_{ANO}^2 = (F_{PA}^+ - F_{PA}^-)^2$$

(Eq. 27)

Where F_{ANO} is the anomalous signal, and F^+ and F^- are the Friedel pairs.

Once the position of the anomalous scatter is identified a Harker diagram can again be constructed to determine the phase of the protein.

One problem with anomalous scattering is the difference between Friedel pairs is small (Rhodes, 2010). This means the phase information determined by the anomalous diffraction may not be enough to determine a suitable electron density map. It may therefore require another derivative or a combination of other phasing methods to determine phases with sufficient accuracy. Alternatively, to increase the phasing power of an anomalous signal it is possible to collect datasets at multiple wavelengths. This method is called multi-wavelength anomalous dispersion (MAD) (Walsh et al., 1999).

3.4.4 Molecular replacement

The most common method for determining phase information is called molecular replacement (Rhodes, 2010). For this technique a solved structure needs to exist which is a structural homologue. The structure factors from this model are then used as an initial estimate for phase information of the new protein.

There are two procedures for molecular replacement depending if the search model is isomorphous to the new protein. If the new protein is isomorphous then the phases of the search model can be used directly to calculate initial estimates for the new structure factors. The model is then refined to improve the phase information.

The more challenging task is if the search model and the new protein are not isomorphous. To determine phase information in this case, the orientation and position of the search model needs to identified in the unit cell. Once the location of the new protein is discovered new structure factors can be calculated.

To place a molecule in the unit cell both the position and orientation need to be found. It should, therefore, be possible to search every location and orientation of the model to find the correct location. However, to simplify the above problem the search can be split into two stages. Firstly, the orientation of the search model is determined, and then the position of search model in the unit cell is calculated. Phaser is the computer programme, usually, used to carry out molecular replacement (McCoy et al., 2007). To determine both the orientation and position of the search model Phaser uses maximum-likelihood calculations in reciprocal space (Evans and McCoy, 2007).

Older methods of molecular replacement used a Patterson function to determine the orientation of the search model; as a Patterson map is independent of the models position in the unit cell but depends on its orientation. Matching the Patterson maps gives an estimate of the orientation of the protein in the unit cell. Once the orientation is determined the position of the protein in the unit cell can be identified by calculating estimates of the structure factors, and then measuring how they compare with the observed structure factors.

3.5 Refinement

The initial phases obtained from the methods described above are estimates and produce an electron density map which is usually difficult to interpret. There are multiple methods to improve phases, which make the map more interpretable. This

is an iterative process which bootstraps the initial phase estimates to phases which are more accurate. These methods include refinement in both the real and reciprocal space.

The initial phase estimates in each Fourier sum vary in reliability as some phases are more accurate than others. To take this into account a weighting factor between 0-1 (*w*) is given to each term in the Fourier series so that:

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} w_{hkl}|F_{hkl}|e^{-2\pi i(hx+ky+lz-\alpha_{hkl})}$$

(Eq. 28)

For a phase which has a low accuracy the weighting factor is close to zero. However, once phases are improved the weighting factors changes to a higher number, which means it will have a greater impact on the calculation of the Fourier sum. After multiple rounds of refinement and improvement in phases *ρ(x, y, z)* will become more accurate and interpretable (Rhodes, 2010).

The process of refinement involves making a change to the map or model, and then the structure factors are re-calculated using this new information. Next the phases from these structures are used with the observed reflection intensities to re-calculate *ρ (x, y, z)*.

The methods available for real space refinement will firstly be discussed such as solvent flattening, which is a process to improve solvent to protein boundaries; non-crystallographic symmetry (NCS); and model building, which involves building and editing the protein structure at an atomic level. Then methods of reciprocal space refinement will be described with a final discussion on the statistical methods used to monitor the refinement.

3.5.1 Real space refinement

Solvent flattening is a method to improve phases by adding additional constraints by assuming in real space the solvent density is flat (Rhodes, 2010). To carry out this technique the map is initially divided into equally sized grids and the average electron

density of this area is calculated. This area is then given a value of zero if the overall average is negative or if the electron density is positive the area is given the value corresponding to the average. This should improve the map by removing any uneven density and help define the protein envelope. The structure factors are then re-calculated and the phases used with the observed reflections to re-calculate the map.

NCS averaging takes advantage of addition symmetry present in the unit cell due to the protein, which cannot be extended over the whole crystal (Murshudov et al., 2011). This usually occurs if in the asymmetric unit there are two protein molecules related by a symmetry operator such as a two-fold rotation. The operator function can be identified by a number of methods. For instance, if the phases were determined by the heavy atom method it is possible to calculate the symmetry operator between groups of heavy atoms. Once the symmetry operator is found in the unit cell, the map can be averaged between the additional symmetry elements. In addition, restraints can be placed on certain phases in reciprocal space to take into account the NCS operator.

Once the phases are improved using the above two methods it should be possible to observe clear protein density in the electron density map. This could be in the form of secondary structure elements such as α helices or in some cases amino acid sides chains. If molecular elements cannot be identified it suggests additional phase information is required (Rhodes, 2010).

In the atomic model building stage a poly-alanine chain is firstly built, and where feasible side chains are added. Bonds restraints are added throughout the refinement as the geometry of amino acids and their side chains have already been determined. Once a round of rebuilding has occurred new structure factors are calculated based on the co-ordinates of the atoms and their scattering factor (f_j). The new phases are used in combination with the observed intensities to calculate a new map, and this map is then used to carry out an additional round of rebuilding. This process is repeated until the phases can no longer be improved.

During the atomic building stage, in the refinement process, the model can easily become biased toward to the atomic co-ordinates as they contribute an increasingly

larger proportion to the phase information. To remove any bias during the atomic building process the following two maps are plotted (Rhodes, 2010):

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}(|F_o| - |F_c|e^{-2\pi i(hx+ky+lz-\alpha_{hkl})}$$

(Eq. 29)

and:

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}(2|F_o| - |F_c|e^{-2\pi i(hx+ky+lz-\alpha_{hkl})}$$

(Eq. 30)

Where F_o are the observed intensities and F_c are the calculated intensity.

Equation 29 and 30 represents a F_o-Fc and $2F_o$-Fc maps respectively. A $2F_o$-Fc map is the normal map that the proteins are built into. This map helps remove bias as it takes into account both the observed and calculated intensities. The F_o-F_c map reveals if there is any significant difference between the atomic model and the observed intensity. If there are extra electrons in an area this will be revealed by positive density in an F_o-F_c map. Alternatively, if an atom is placed in a position with lower electron density this will appear as negative density. Using these two maps it is possible to spot any inaccuracies in the atomic model.

3.5.2 Reciprocal space refinement

In addition to refining the structure in real space it is also possible to carry out phase improvement in reciprocal space. This is because for any atomic model the value of $|F_c|$ should match for $|F_o|$. It is therefore possible to refine the co-ordinates of the protein to produce a better match for F_c. This used to be routinely done using the mathematical method called least square which has the following formula (Murshudov et al., 2011).

$$\phi = \sum_{hkl} w_{hkl}(|F_o| - |F_c|)^2$$

(Eq. 31)

Where w_{hkl} is the waiting factor of each term.

Additional parameters can be added to equation 31 such as B-factors, bond restraints and occupancy, which help to improve the model. Normally, at the start of the refinement these addition parameters act as restraints as they are given pre-set values. This helps prevent the model from getting trapped in an incorrect minimum, but as the model improves these restraints are released, which should help lead to a convergence of $|F_o|$ and $|F_c|$. Modern refinement programmes normally use maximum-likelihood calculations in reciprocal space to decrease the difference between $|F_c|$ and $|F_o|$ (McCoy, 2004).

3.5.3 Monitoring the refinement

Throughout the refinement a range of statistical indexes exist to allow the accuracy of the model to be determined. One of the most common statistical tools is called the residual index (R-factor) (Brünger, 1992). The R-factor takes into account the difference between the observed and calculated intensities in the following equation:

$$R = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|}$$

(Eq. 32)

For a model where the observed intensities are in agreement with the calculated intensities the R-factor will be small and should approach zero. However, an R-factor of almost 0.5 represents a random distribution of atoms in a unit cell. For a model which has a resolution of 2 Å, it should be possible to obtain an R-factor of below 0.2.

At the start of refinement 5 % of reflections are randomly taken as a free R-set (Brünger, 1992). This free R-set is not used to calculate the structure factors and is not refined against during model building. Comparing the calculated reflections against the free R set should give an unbiased analysis on how accurate the model is at describing the observed reflections.

Apart from the R-factors there are a number of other parameters which can help assess how accurate a model is which include the protein geometry, bond angles and steric clashes (Rhodes, 2010).

In this chapter I have described how to obtain crystals from aqueous protein solutions and then optimise them to produce crystals that diffract strongly. Then crystal symmetry was discussed; along with how this relates to the diffraction pattern of a crystal. Finally, the method of obtaining an electron density map from the diffraction pattern was described.

Chapter 4. Structure of Ctf18^{C}-Dcc1-Ctf8

4.1 Summary on the Function of RFC^{Ctf18}

The RFC^{Ctf18} complex is made from the seven subunits Ctf18, Rfc2-5, Dcc1 and Ctf8 (Mayer et al., 2001). It was first discovered due to its genetic roles in the establishment of sister-chromatid cohesion, and more recently has been shown to effect the DRC (Crabbé et al., 2010; Pan et al., 2006). RFC^{Ctf18} belongs to a class of clamp loading or unloading complexes known as RFC like complexes. These complexes all contain the four subunits Rfc2-5, and are defined by their large subunit (Kubota et al., 2014), which in the case of RFC^{Ctf18} is Ctf18. There are four RFC like complexes encoded in the eukaryotic genome, with RFC^{Ctf18} being the only complex formed from seven subunits (Mayer et al., 2001). Work has demonstrated that *in vitro* RFC^{Ctf18} can both load and unload PCNA from a range of DNA substrates; however, it was suggested that the main *in vivo* target was as a PCNA unloader (Bylund and Burgers, 2005). This was different to the conclusion drawn by ChIP analysis, which rationalised that RFC^{Ctf18} was a PCNA loader as in Ctf18 deletion strains the amount of PCNA on chromatin was reduced (Lengronne et al., 2006). Furthermore, *in vitro*, it was discovered that the two subunits, Dcc1 and Ctf8, took no part in the unloading or loading activity of PCNA (Bylund and Burgers, 2005). It is interesting to note that deleting either Dcc1 or Ctf8 has the same phenotype in yeast as a *ctf18Δ* strain in both the DRC and sister-chromatid cohesion. Therefore, I hypothesised that these subunits must play a regulatory role on the complex. To gain an insight into the possible function of the Dcc1-Ctf8 sub-complex I decided to obtain the crystal structure of this heterodimer. It was hoped that I would then be able to rationalise how Dcc1 and Ctf8 regulate the RFC^{Ctf18} complex.

4.2 Expression and Purification of Dcc1-Ctf8

The expression of the full length Dcc1-Ctf8 sub-complex was carried out in *E. coli* by the protocol presented in Section 2.2. The sub-complex expressed well and could be purified to a high standard using a three-step purification protocol (Figure 4.1A). The resulting protein was concentrated, and crystallisation trials were attempted at 10

mgmL^{-1} and 20 mgmL^{-1} in a range of commercial available screens. Crystals were obtained in a range of conditions and were optimised by varying the pH and precipitant concentration. Unfortunately, after multiple rounds of optimisation the crystal diffraction could not be improved beyond ~20 Å; therefore, I decide the construct needed to be improved. To do this a range of sequence alignments for both subunits were carried out to see if there appeared to be any conserved domains; unfortunately, the analysis did not reveal any clear results (data not shown). Consequently, limited proteolysis was attempted.

Limited proteolysis uses a 'small' amount of protease to cut a protein at flexible or unstable regions. It is hoped that the identification of these boundaries will enable the design of new stable constructs that may crystallise and diffract to a higher resolution. The results from limited proteolysis conducted on the Dcc1-Ctf8 sub-complex revealed that Dcc1 and Ctf8 both degraded to a stable construct (Figure 4.1B). The resulting bands were analysed by both Edman N-terminal sequencing and in-tact mass by MALDI (this work was performed by the PNAC facility at the University of Cambridge). The data indicated that the proteases had cleaved Dcc1 and Ctf8 at residue 90 and 18 respectively. As a result, these construct boundaries were re-cloned and expressed in *E. coli*. The purification of Dcc1^{FL} and Ctf8$^{18\text{-}133}$ resulted in the formation of a stable complex which purified to a high quality (Figure 4.1C). However, when the expression of Dcc1$^{90\text{-}380}$ and Ctf8$^{18\text{-}133}$ was attempted the proteins could not be purified; suggesting that the N-terminal region of Dcc1 was required for the interaction with Ctf8.

The complex formed from Dcc1^{FL} and Ctf8$^{18\text{-}133}$ was purified and placed into crystal trials at 15 mgmL^{-1} using a variety of commercially available crystal screens. Potential hits were optimised by changing the pH and concentration of the precipitant. Unfortunately, even after multiple rounds of crystal optimisation the diffraction could not be improved beyond a resolution of 3.5 Å.

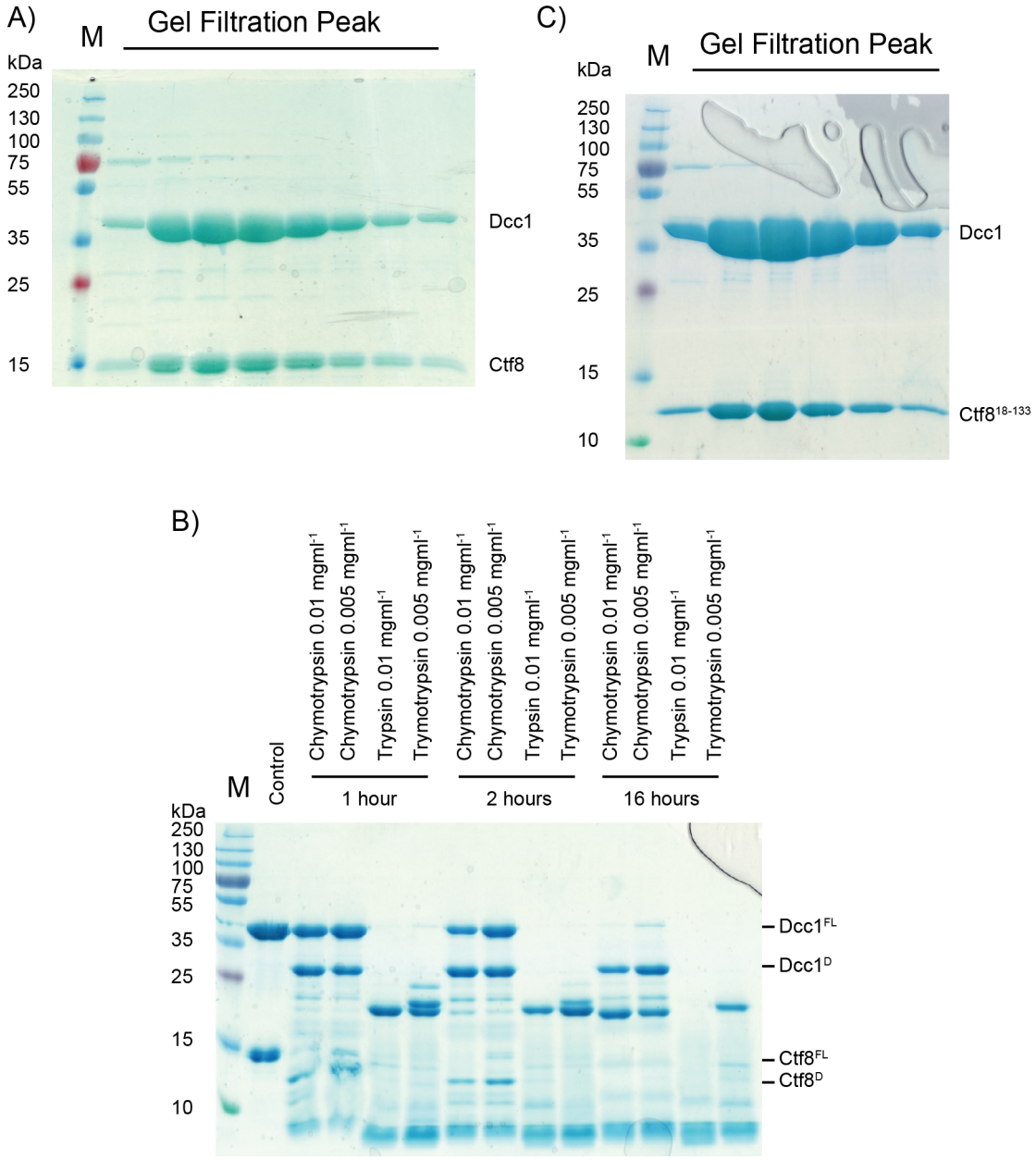


Figure 4.1. Construct Design for the Dcc1-Ctf8 Sub-Complex.

A) Purification of the full length Dcc1-Ctf8 sub-complex, the SDS-PAGE presented is of the final purification step. B) Limited proteolysis carried out on the Dcc1-Ctf8 sub-complex. Degraded bands labelled 'D' were analysed by MS to determine new construct boundaries. C) Purification of the sub-complex of Dcc1^{FL} and Ctf8$^{18-133}$, the SDS-PAGE presented is of the final purification step.

4.3 Crystallisation of the C-Terminus of Dcc1

As a result of the poor diffraction of the Dcc1-Ctf8 crystals, Dcc1$^{90-380}$ was expressed on its own. The single protein expressed well from *E. coli* and purified to a mono-species after a three-step procedure (Figure 4.2A). Crystallisation was attempted in a range of commercially available screens at 15 mgmL^{-1}. The construct crystallised as a hollow rod shape in a variety of conditions, growing to a maximum length of ~900 μm in 0.2 M potassium sodium tartrate and 18 % PEG 3350 (Figure 4.2B). These crystals were harvested after two days in liquid nitrogen with the addition of 20 % ethylene glycol. The largest crystal diffracted to a resolution of 2 Å at Diamond Light Source on beamline I04-1. To determine a usable electron density map, the phase problem needed to be solved. The phases can usually be determined by either molecular replacement or experimental phasing. However, as there is no known structural homologue for Dcc1 the latter method has to be used.

To determine the phases for Dcc1$^{90-380}$, protein containing only selenomethionine was obtained and crystallised. To do this, Dcc1$^{90-380}$ was expressed in *E. coli* in minimal media where selenomethionine was the only source of methionine. This protein was then purified in the same three-step procedure as the native protein. Crystals trays were set up at 15 mgmL^{-1} in an optimisation grid screen from the condition that produced the best dataset for the native crystals. Selenomethionine crystals grew as large plates to a length of ~700 μm (Figure 4.2C) and were harvested in the same way as the native crystals. The crystals were shot at Diamond Light Source on beamline I03 and a two wavelength multi-wavelength anomalous diffraction (MAD) dataset was collected at the peak (0.9796 Å) and inflection (0.9797 Å) wavelengths to a resolution of 3 Å (Figure 4.2D).

From the MAD dataset the AutoSHARP package (Bricogne et al., 2003) was able to determine the position of 33 out of 36 selenomethionines in the unit cell. The initial estimate of 36 selenomethionines was calculated based on 4 molecules in the asymmetric unit. This was determined using the Matthews coefficient, which gave a probability of 0.65 for 4 molecules (V_m = 2.15 Å^{3}Da^{-1}) and a percentage solvent of 41.77 % (Matthews, 1968). From the initial phases a map was generated and refined using standard procedures such as solvent flattening and non-crystallographic

symmetry (NCS) averaging. NCS averaging was carried out based on the Mathews coefficient of 4 molecules being in the asymmetric unit. Buccaneer was then used to trace an initial model (Cowtan, 2006), and multiple round of refinement and rebuilding were carried out in Coot (Emsley and Cowtan, 2004) and Refmac5 (Murshudov et al., 2011) respectively. This structure was then used as a model for molecular replacement in Phaser for the high resolution native dataset (McCoy et al., 2007). After which, further rebuilding and refinement were conducted. Data collection and final refinement statistics are given in Table 2.5. The final structure of Dcc1$^{90-380}$ is shown in Figure 4.2E and is discussed in detail in the next sections.

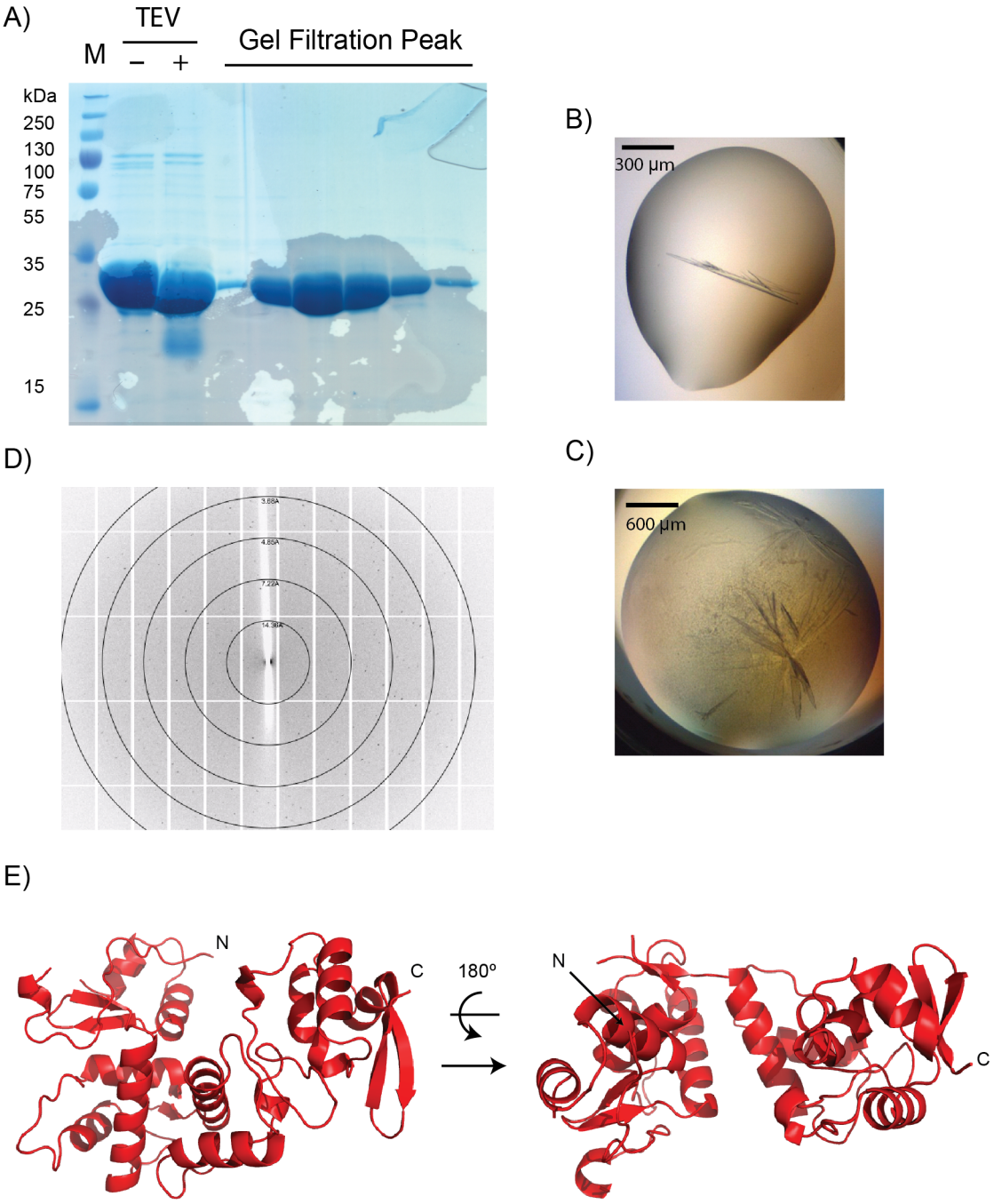


Figure 4.2. Crystallisation of Dcc1$^{90-380}$.

A) The final step in purification of Dcc1$^{90-380}$ analysed by SDS-PAGE. B) Crystallisation of native Dcc1$^{90-380}$. C) Crystallisation of selenomethionine Dcc1$^{90-380}$. D) Diffraction pattern of native Dcc1$^{90-380}$ crystals. E) Overall crystal structure of Dcc1$^{90-380}$.

4.4 Dcc1 Contains Three Tandem Winged Helix Domains

The role of Dcc1 in the RFC^{Ctf18} complex is currently unknown as it takes no part in the loading or unloading of PCNA (Bylund and Burgers, 2005). However, when Dcc1 is mutated it shows the same phenotype as Ctf18 deletion strains. To determine if the C-terminus of Dcc1 was similar to any solved structures the co-ordinates of Dcc1 were uploaded to the DALI server (Holm and Rosenström, 2010). The DALI server uses an algorithm that compares all known structures against the co-ordinates uploaded; it was hoped that uncovering structural homologues could help identify a function for the Dcc1 subunit.

The results from the DALI server unexpectedly revealed that Dcc1 contained three tandem WH domains (Figure 4.3A and B). These domains have been called WH1, WH2 and WH3 which have the residue boundaries of 182-252, 253-318 and 319-380 respectively. They were identified by having a high Z-score with WH containing proteins for instance WH1, WH2 and WH3 had scores of 5.7, 7.4 and 4.9 with PKZ (PDB ID: 4KMF), Cullin-1 (PDB ID: 3TDU) and DsrD (PDB ID: 1WQ2) respectively (Figure 4.3C).

WH domains usually contain the secondary structure of α1-β1-α2-α3-β2-β3 (Harami et al., 2013). All of the identified WH domains have this morphology; however, in WH2 α3, β2 and β3 are partially disordered. This indicates they could be mobile or unstable in the crystal lattice, which could possible allow the C-terminus some flexibility permitting WH3 to exist in multiple geometries.

The discovery that Dcc1 has multiple WH domains gives an indication for the potential role of this subunit. WH domains structurally have been conserved throughout evolution and perform a variety of functions throughout the cell (Harami et al., 2013). Many proteins use WH domains to interact with DNA substrates. For instance, the transcription factor HNF3γ binds to dsDNA through the secondary structure elements α3, β2 and β3 of its WH domain. The use of these three structural features is generally conserved in the WH-DNA interactions (Harami et al., 2013). Other DNA binding WH domains can interact with a variety of other nucleotide substrates in either a sequence dependent or independent manner such as Z-DNA,

RNA hairpin structures, ssRNA and ssDNA. Alternatively, other WH domains are involved in protein-protein interactions like the cohesin subunit Scc1, which binds to Smc1 via a WH domain present at its C-terminus (Haering et al., 2004). From the above discussion it is probably that WH1, WH2 and WH3 participate in either protein-protein interaction through binding to Ctf8, Ctf18 or an unknown protein partner, or Dcc1 could bind to a specific DNA substrate (See Chapter 5).

The N-terminal region of Dcc1 in the crystal structure from residue 114-182 is folded and most likely acts as a linker to another domain. This part of Dcc1 did not show any clear results from the DALI server search. The residues 90-113 were unstructured in this crystal suggesting they could be unstable without the rest of the N-terminal region of Dcc1 or require the binding of another protein to form a stable structure. Furthermore, the results that Dcc1 was cleaved at residue 90 indicates that this part of Dcc1 is flexible.

The determination of the structure of the C-terminal region of Dcc1 means a model is now available which could be used in molecular replacement for the 3.5 Å dataset collected on the Dcc1-Ctf8$^{18-133}$ crystals. However, due to the flexibility of both the N-terminal region of Dcc1 and Ctf8 I was unable to determine any additional information on the interaction between Dcc1 and Ctf8.

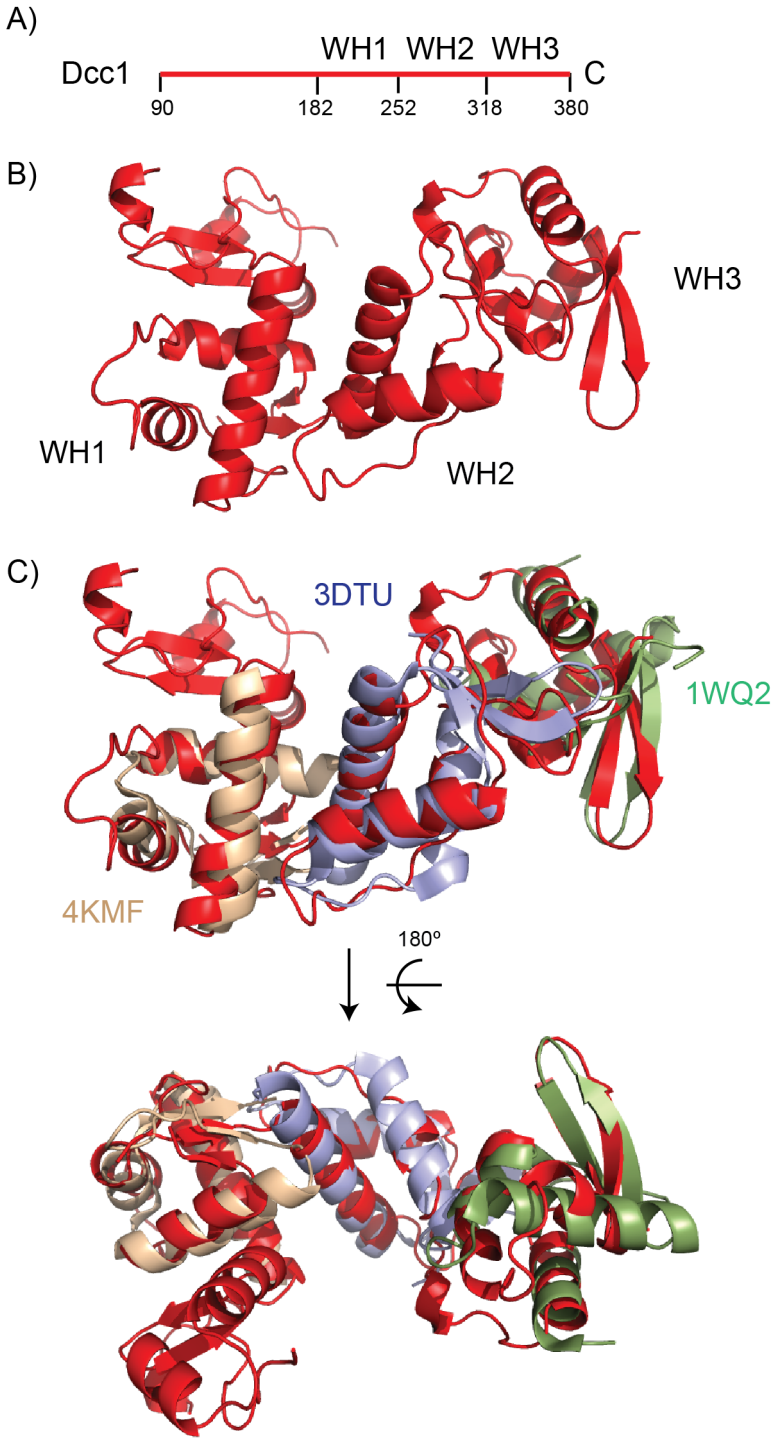


Figure 4.3. Organisation of the Three WH Domains of Dcc1.

A) *Domain organisation of Dcc1. B) The localisation of the three WH domains on Dcc1. C) Structural alignments of the three WH domain containing proteins PKZ (PDB ID: 4KMF, orange), Cullin-1 (PDB ID: 3TDU, blue) and DsrD (PDB ID: 1WQ2, green).*

4.5 Crystallisation of Ctf18^{C}-Dcc1-Ctf8

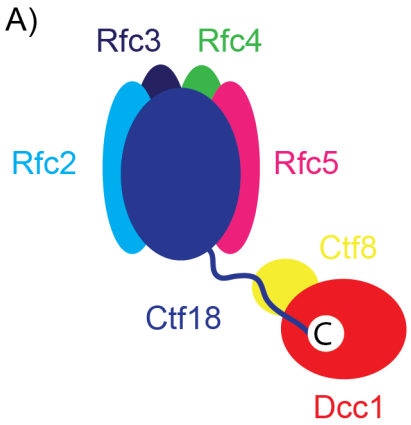
To further investigate the Dcc1-Ctf8 sub-complex I firstly wanted to discover the structure of Ctf8 and secondly look at how the complex interacts with Ctf18. The co-expression of the full RFC^{Ctf18} complex has been previously carried out in budding yeast (Bylund and Burgers, 2005), and the authors were able to show that deletion of the last 80 residues of Ctf18 was enough to disrupt binding of the 5 subunit complex to the Dcc1-Ctf8 sub-complex. Therefore, it was suggested that the C-terminus of Ctf18 acts a bridge between the core RFC^{Ctf18} complex and the Dcc1-Ctf8 sub-complex (Figure 4.4A). To fulfil both of the above aims I decided to form a heterotrimer through co-expressing Ctf18, Dcc1 and Ctf8. It was hoped this would result in a well expressed complex which could stabilise Ctf8 and the N-terminal region of Dcc1. The crystallisation of this heterotrimer would further provide information on the interaction between Ctf18 and the Dcc1-Ctf8 sub-complex.

To design a stable complex a sequence alignment was carried out of Ctf18 from five organisms (Figure 4.4B). The results of this alignment indicated there was limited sequence conservation over the whole C-terminal region. However, the extreme C-terminus, from residue 717 to 741, had high sequence conservation. As a result of this alignment, the three constructs Ctf18$^{598-741}$, Ctf18$^{539-741}$ and Ctf18$^{666-741}$ were cloned into pET-28a with the addition of a His_6 tag and TEV protease site. The vectors were transformed into *E. coli* for co-expression and purification with Dcc1 and Ctf8 using the protocol presented in section 2.3.

From the three Ctf18 constructs cloned only Ctf18$^{666-741}$ expressed, indicating that the longer constructs were insoluble in solution, possible because they require the N-terminal region of Ctf18 to be folded correctly. Ctf18$^{666-741}$ formed a heterotrimer in solution with Dcc1 and Ctf8, which was stable throughout the three-step purification procedure. The Ctf18$^{666-741}$-Dcc1-Ctf8 complex (hereafter called Ctf18^{c}-Dcc1-Ctf8) could be purified to a high standard and expressed at a level sufficient to be placed into crystal trials (Figure 4.5A). The protein was concentrated to 15 mgmL^{-1} and crystal trays were set up in a variety of commercially available screens. An initial hit in 0.1 M Bis-Tris propane pH 6.5, 0.2 M NaBr and 25 % PEG 3350 was identified, which diffracted to ~4 Å. To improve the resolution crystals were optimised by varying

the pH and precipitant concentration, which produced slightly better crystals. However, to further improve crystals quality larger drops of 4 μL were used along with an increased protein concentration of 35 mgmL^{-1}. These crystals grew as thick plates to a size of ~700 μm (Figure 4.5B). The crystals were harvested after 10 days in 15 % ethylene glycol and stored in liquid nitrogen. Data was collected at Diamond Light Source on beamline I03, and the best crystal diffracted to 2.3 Å (Figure 4.5C).

To solve the structure of Ctf18^{c}-Dcc1-Ctf8 molecular replacement was carried out in Phaser using Dcc1$^{90-380}$ as a starting model (McCoy et al., 2007). AutoBuild (Adams et al., 2010) was used to trace an initial model after which, multiple rounds of rebuilding and refinement were carried out in Coot (Emsley and Cowtan, 2004) and Refmac5 (Murshudov et al., 2011) respectively. Full refinement and data collection statistics are shown in Table 2.6.



B) Ctf18 C-terminal

```
S.cerevisiae KSLIFEILPYLDSMISSDFNKIRNLKLK---QAIMEELVQLLKSFQLNLIQNRSEGFDVR 601
S.pombe      RSILLELIRTILITINPTLKQKEDSMPRSALPSKIEHAINILNHYNLRFQQLPV--GDGN 833
X.laevis     QSLVLDALCLLLDVLSPKLRPVNTQLFSTKEKQQLAELINTMLAYNLTYHQERT--MEGQ 838
M.musculus   QALVLDTLCLLLDVLAPKLRPVSTQLYSAHEKQQLSCLVGTMLAYSLTYHQERT--PDGQ 806
H.sapiens    QALLLDALCLLLDILAPKLRPVSTQLYSTREKQQLASLVGTMLAYSLTYRQERT--PDGQ 812
             :::::: :  :   :   :.             :   :  :  :.*   *      : .

S.cerevisiae GGLTIDPPIDEVVLLNPKH-----INEVQHKRANNL-SSLLAKIEENRAKKRHIDQVTED 655
S.pombe      YVYRLEPPLDELVWDAPTS-----SYSVRQMLSQELLKKRLADIKK-----QT---LTDN 880
X.laevis     YVYKLDPNVEDVCRFPDLPVRKPLTYQTKQLIAREIELERMRRTEAF-QQARNAGRDNTT 897
M.musculus   YLYKLEPNVEEVCRFPELPARKPLTYQAKQLIAREIEMEKMRRAEAL-AWARSGPQVDQG 865
H.sapiens    YIYRLEPNVEELCRFPELPARKPLTYQTKQLIAREIEVEKMRRAEAS-ARVENSPQVDGS 871
                 ::* ::::             ..::  :.::  . :   :      .

S.cerevisiae RL--------------------QSQEMHSKKVKTGLNSS-SSTIDFFKNQYGLLKQTQE 693
S.pombe      PT--------------------SS---NSSRKRKDFNSGKAIKRDFFGR----IISEPK 912
X.laevis     AAAAVKTADPKGAKSAAKPAALNHEQRLEN--IMKKATFEEKPEKDFFGRQIVKKVAAPV 955
M.musculus   SSGPASLWTDSGEKGTRQPAPRNHEQRLEH--IMKRATVQEQPERDFFGRVVIRKVAVPS 923
H.sapiens    PPGLEGLLGGIGEKGVHRPAPRNHEQRLEH--IMRRAAREEQPEKDFFGRVVVRSTAVPS 929
                                 .   .               *** .

S.cerevisiae LEETQKTIGSDETNQADDCNQTVKIWVKYNEGFSNAVRKNVTWNNLWE- 741
S.pombe      SEAVTSN-NAALNTGDHPVSIKHAINIKFHDGFSNAVRKPISLNEILNF 960
X.laevis     TA-SANQEESVE-RRIGKAVGNSDVWFRFNEGVSNAVRRNIYIKDLL-- 1000
M.musculus   REVEAPQKDADE-WRMGVAVGRSEVWFRFNEGVSNAVRRSLYIRDLL-- 969
H.sapiens    AGDTAPEQDSVE-RRMGTAVGRSEVWFRFNEGVSNAVRRSLYIRDLL-- 975
                      :             : .::.:*.*****: :  .::
```

Figure 4.4. Construct Determination for Ctf18-Dcc1-Ctf8.

A) Cartoon representation of the structure of RFC^{Ctf18}, Ctf18 (dark blue), Ctf8 (yellow) and Dcc1 (Red). B) Sequence alignment of the C-terminus of Ctf18 from yeast to humans. Arrows indicate the construct design for expression with the Dcc1-Ctf8 sub-complex.

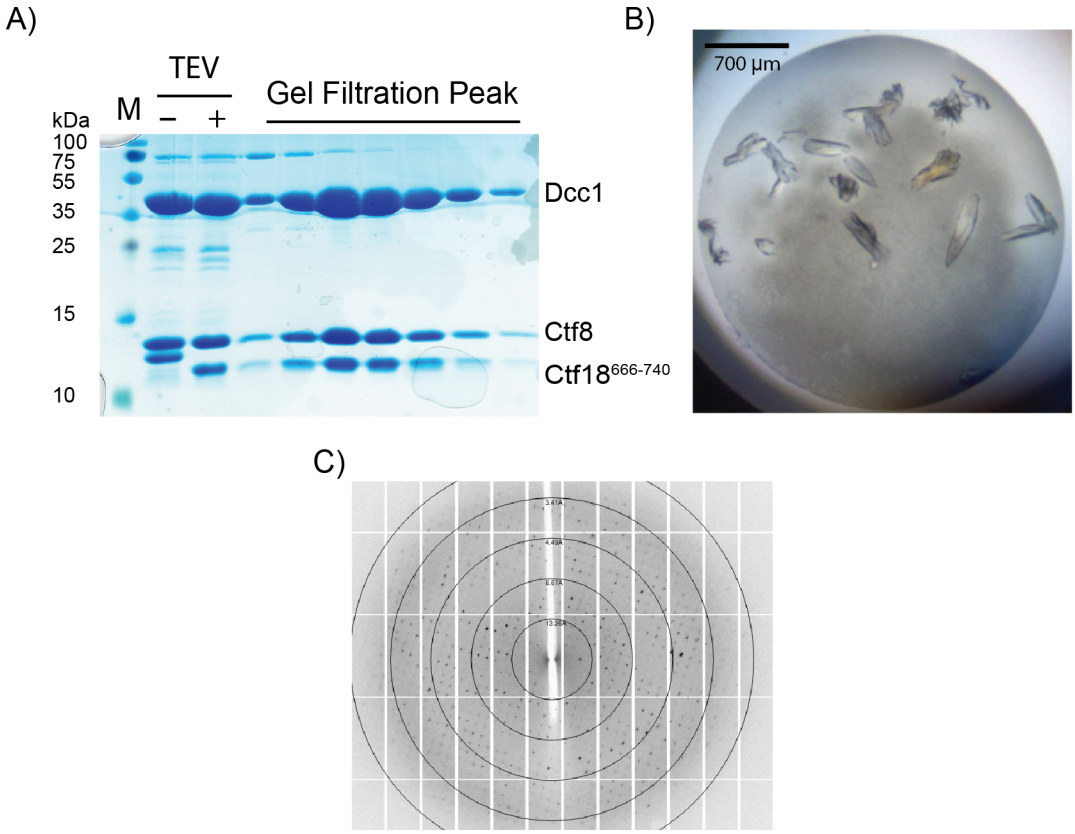


Figure 4.5. Purification and Crystallisation of Ctf18^{C}-Dcc1-Ctf8.

A) Purification of Ctf18^{c}-Dcc1-Ctf8. The SDS-PAGE shows the final step in the three-step purification procedure. B) Optimised crystals of Ctf18^{c}-Dcc1-Ctf8. C) Diffraction from optimised native crystals of Ctf18^{c}-Dcc1-Ctf8.

4.6 Overall Structure of Ctf18^{c}-Dcc1-Ctf8

In the Ctf18^{c}-Dcc1-Ctf8 heterotrimer the C-terminus of Dcc1 folds back towards its N-terminus making the structure look like a 'hook'. The interaction domain consists of the N-terminus of Dcc1, Ctf8 and Ctf18^{c} (Figure 4.6). The N-terminus of Dcc1 and Ctf8 make a large dimer interface which is formed from 18 intertwined β-strands and two α-helices. The heterodimer interface is well conserved (see Figure 4.8 in section 4.7) with the formation of multiple antiparallel β-strands between Dcc1 and Ctf8. The number of interactions between Ctf8 and Dcc1 make it unlikely that Ctf8 can exist by itself in cells, and Dcc1 is probably required for the correct folding of Ctf8 both *in vitro* and *in vivo*.

Ctf18^{c} binds to the Dcc1-Ctf8 heterodimer interface as an extended peptide, forming an N-terminal antiparallel β-strand with Ctf8, and a partially buried C-terminal helix. This helix binds to a β-barrel like structure formed from antiparallel β-strands from both Dcc1 and Ctf8 (Figure 4.6). Interestingly, crystals formed from full length Dcc1 and Ctf8 did not diffract strongly, indicating that the heterodimer is stabilised by Ctf18^{c}. Residues 666-713 of Ctf18^{c} are unstructured in this structure, suggesting they form a flexible linker to the RFC like domain of Ctf18. The C-terminus of Dcc1 is found in the same conformation as the Dcc1$^{90\text{-}380}$ structure, and is made predominantly from α-helices.

To determine if the dimerisation motif found in the Dcc1-Ctf8 sub-complex has any structural homologs the co-ordinates of the heterotrimer were uploaded to the DALI server (Holm and Rosenström, 2010). The results indicated that the heterodimer interface is similar to four other proteins, which fold to form a 'triple' β-barrel structure. The four complexes identified were the three transcription factors Rap30/74 (Gaiser et al., 2000), Sfc1/7 (Taylor et al., 2013) and A49/34.5 (Geiger et al., 2010); as well as the trimer RNase H2 (Shaban et al., 2010) (Figure 4.7). The 'triple' β-barrel fold seems to be used as an architectural domain linking different domains of larger complexes. In addition, the Ctf18^{c}-Dcc1-Ctf8 heterotrimer has the same modular organisation to the above transcription factors as they have this domain followed by

WH domains, which raises the possibility that the Dcc1-Ctf8 sub-complex evolved from a distant relative of one of these transcription factors.

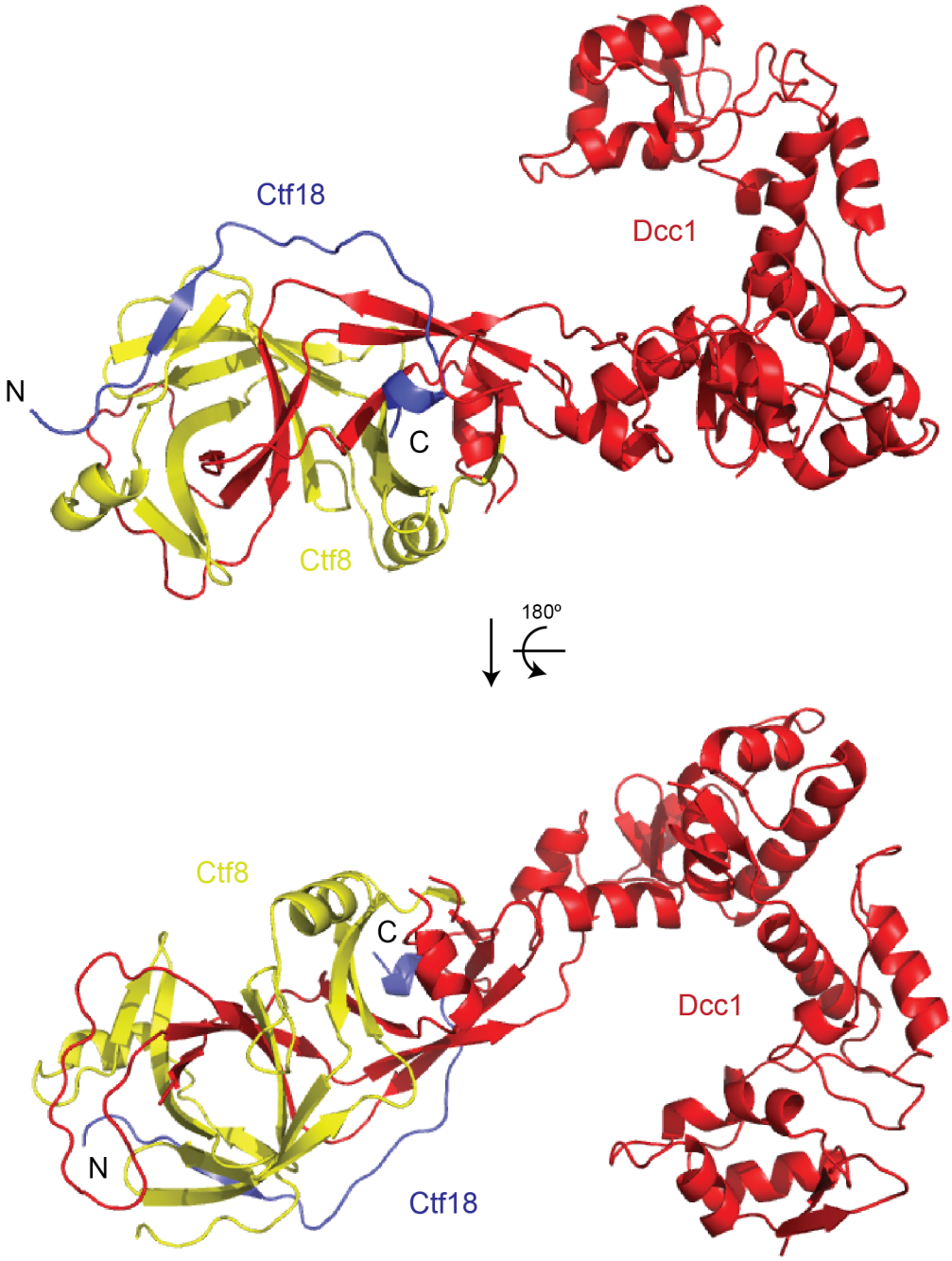


Figure 4.6. Structure of the Ctf18^{c}-Dcc1-Ctf8 Heterotrimer.

Ctf18 (dark blue), Ctf8 (yellow) and Dcc1 (red). The N- and C-terminus of the extended peptide of Ctf18^{c} are indicated with 'N' and 'C' respectively.

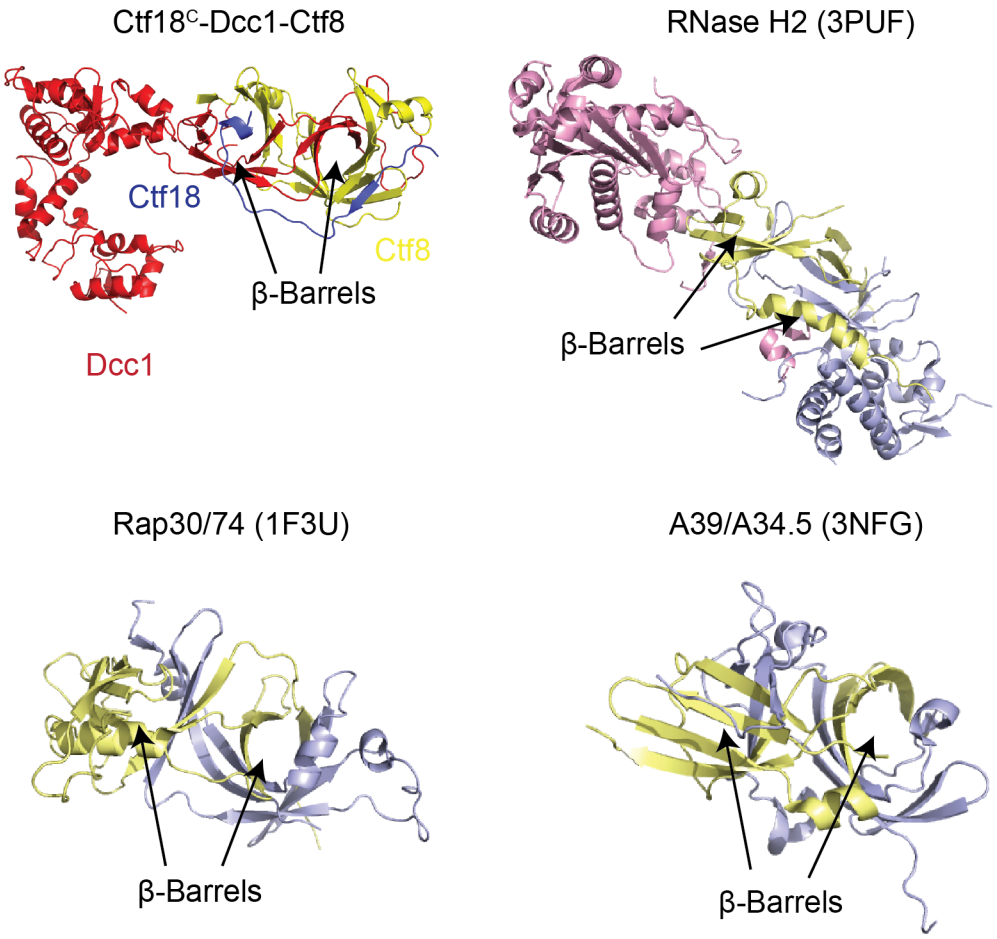


Figure 4.7. Comparison of Known 'Triple' β-Barrel Complexes.

Two of the three β-barrels are indicated in each complex. RNase H2 is formed from the three subunits RNase H2A (pink), RNase H2B (blue) and RNase H2C (yellow). The Rap30/74 heterodimer contains the subunits Rap30 (blue) and Rap74 (yellow). The A39/A34.5 heterodimer is made from the proteins A39 (yellow) and A34.5 (blue). The PDB codes are indicated in the figure.

4.7 Ctf18 Binds to Dcc1-Ctf8 as an Extended Peptide

The C-terminal region of Ctf18 binds to a conserved groove that runs along the Dcc1-Ctf8 heterodimer (Figure 4.8A), with Ctf18 making multiple contacts with both Dcc1 and Ctf8. The N-terminus of Ctf18^{c} makes an antiparallel β-strand with Ctf8 on the third barrel, with the residue E724 from Ctf18 forming a salt bridge with the conserved R128 of Ctf8. The extreme C-terminal α-helix of Ctf18 is partially buried in the first barrel of the heterodimer (Figure 4.8B). This first barrel forms a hydrophobic pocket, which is made from aromatic residues from both Ctf8 and Dcc1. W736 and W740 of Ctf18 interact with a range of hydrophobic residues from this barrel including the N-terminal proline of Ctf8 which contacts W736 of Ctf18. Due to the importance of these two tryptophans of Ctf18 it is surprising that these residues are poorly conserved in an alignment between yeast to humans (Figure 4.4B). However, on closer inspection these residues are normally hydrophobic, suggesting the overall binding mechanism is likely to similar between all organisms.

Work conducted in budding yeast supports the crystal structure of Ctf18^{C}-Dcc1-Ctf8 as, *in vivo,* mutating both W736 and W740 to alanine causes a dissociation of the Dcc1-Ctf8 sub-complex from RFC^{Ctf18} in co-immunoprecipitation experiments (García-Rodríguez et al., 2015). In addition, these mutant strains have a sensitivity to hydroxyurea and a delayed activation of the DRC; shown by a reduced level of Rad53 phosphorylation, and these strains also have an increased defect in sister-chromatid cohesion. Taken together, the above results indicate that W736 and W740 are crucial for the stable binding of Ctf18 to Dcc1 and Ctf8.

The mechanisms that RNase H2A contacts the RNaseH2B-H2C heterodimer is similar to the interaction between Ctf18^{C} and the Dcc1-Ctf8 sub-complex (Figure 4.7). The C-terminus of RNase H2A binds across the 'triple' β-barrel structure like the extended peptide conformation of Ctf18 (Shaban et al., 2010). However, RNase H2A also makes a range of contacts with the side of the dimerisation domain. This is not observed in the Ctf18^{C}-Dcc1-Ctf8 structure as residues 666-713 were unstructured in this structure.

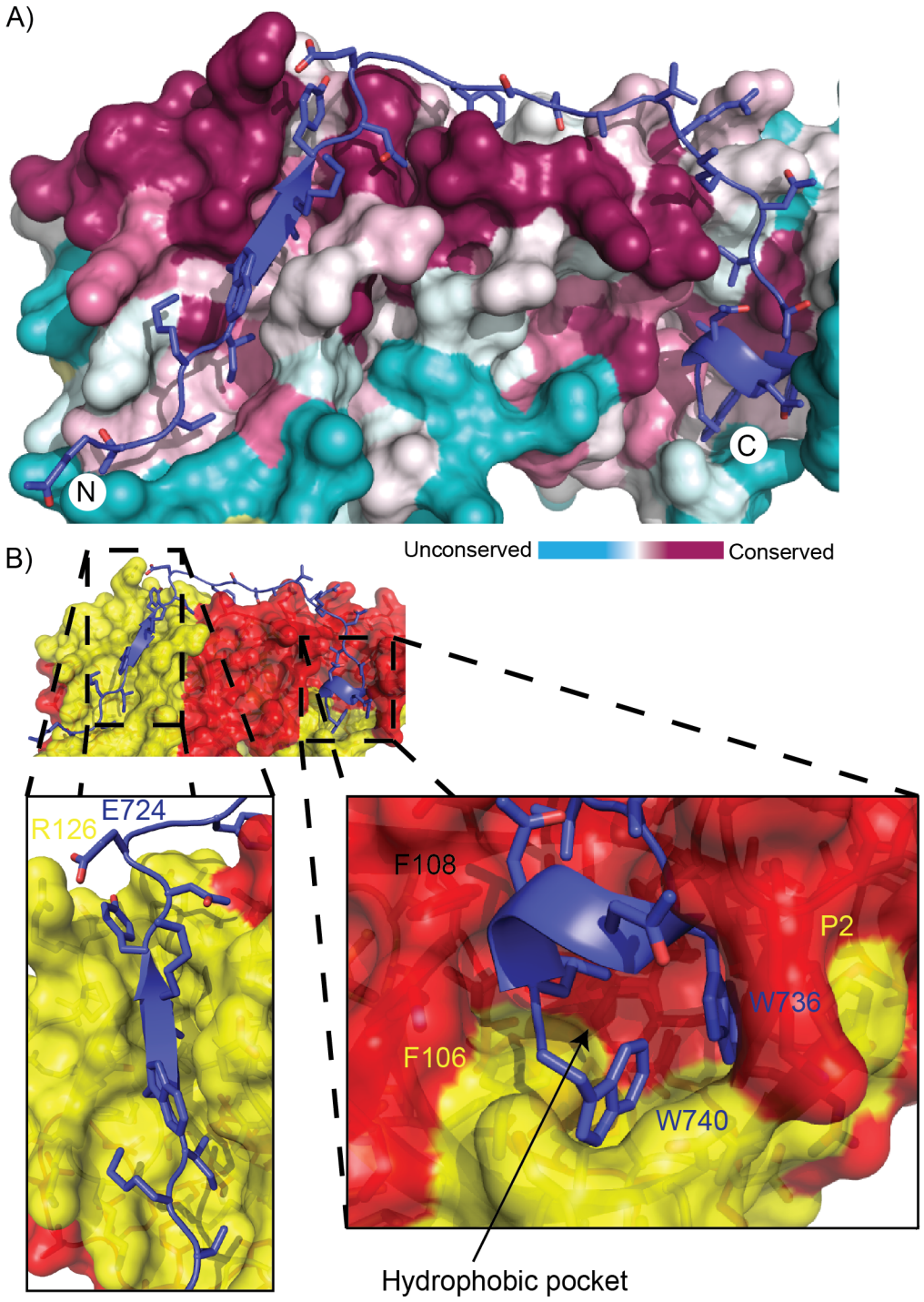


Figure 4.8. Ctf18 Binds to a Conserved Groove on the Dcc1-Ctf8 Heterodimer.

A) Structural conservation of Dcc1 and Ctf8, conserved residues (maroon) and unconserved residues (light green). Ctf18 is shown in dark blue. B) Surface view of the interaction between Ctf18 (dark blue) and Dcc1 (red)-Ctf8 (yellow). Left panel indicates the salt bridge formed from R126 and E724 from Ctf8 and Ctf18 respectively. Right panel shows the partially buried C-terminal helix of Ctf18.

Chapter 5. Biochemical Analysis of $Ctf18^{C}$-Dcc1-Ctf8

5.1 Analysis of the WH Domains of Dcc1

WH domains are usually either involved in protein-protein interactions or bind to a variety of DNA substrates. Due to the location of RFC^{Ctf18} at the replication fork it is likely that Dcc1 could contain additional DNA interacting sites. To determine if the WH domains of Dcc1 can bind DNA the surface conservation and charge were analysed (Figure 5.1). The results indicated that the surface of WH3 and WH2 were highly conserved (Figure 5.1B), and these regions corresponded to a highly basic patch when the surface charge of Dcc1 was studied (Figure 5.1C). Basic residues are usually involved in binding DNA as the backbone of DNA is highly negatively charged. The conserved basic residues found in WH3 were predominantly located on α3, β2 and β3, which is similar to other WH containing proteins that interact with DNA. This further supports the conclusion that Dcc1 can interact with DNA; however, additional biochemical clarification is required.

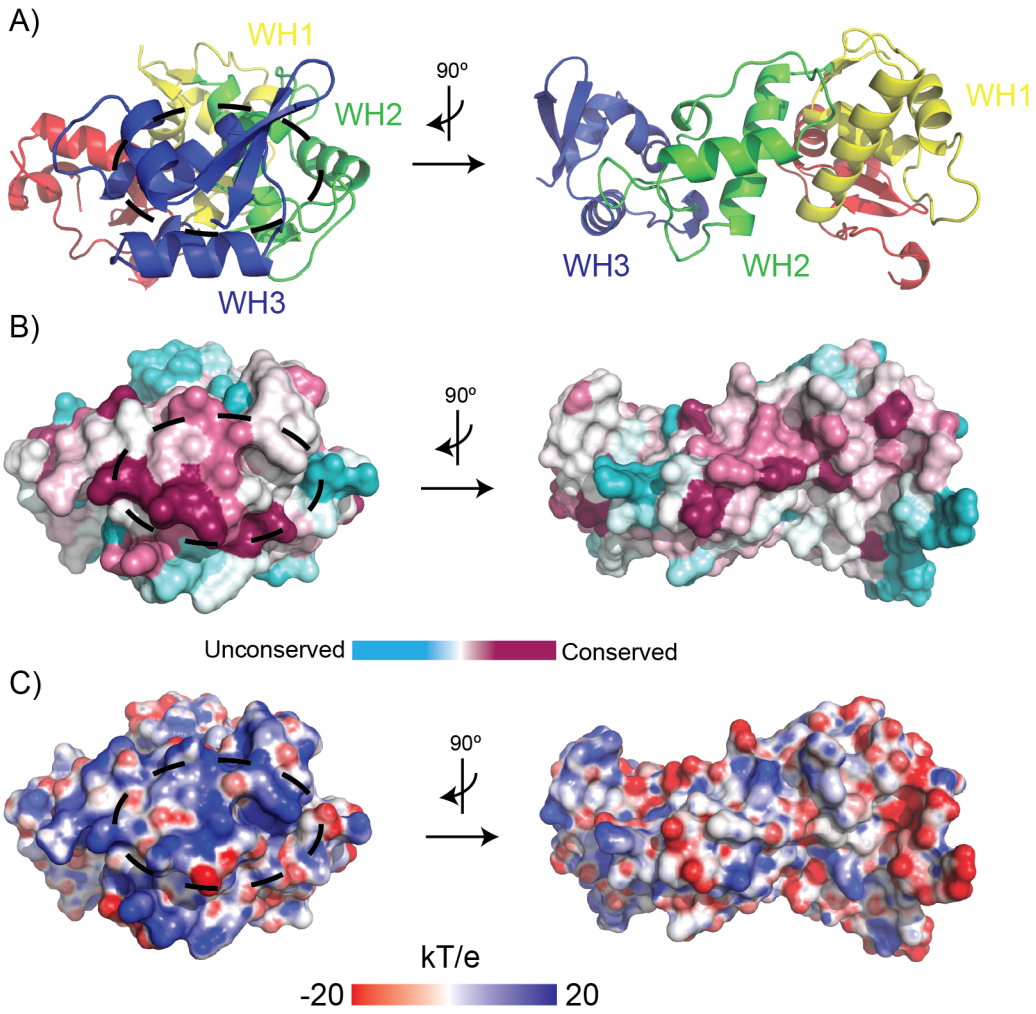


Figure 5.1. Surface Analysis of the WH Domains of Dcc1.

A) Cartoon representation of the C-terminus of Dcc1 coloured as follows; linker region (red), WH1 (yellow), WH2 (green) and WH3 (blue). B) Surface conservation of the C-terminus of Dcc1 in the same orientation as in (A). Conserved and unconserved residue are shown in maroon and light green respectively. C) Surface charge of the C-terminus of Dcc1 in the same orientation as in (A). Basic and acidic areas are shown in blue and red respectively. The dotted cycle indicates a conserved basic patch located on WH3 of Dcc1.

5.2 Dcc1 can Bind to Both ssDNA and dsDNA

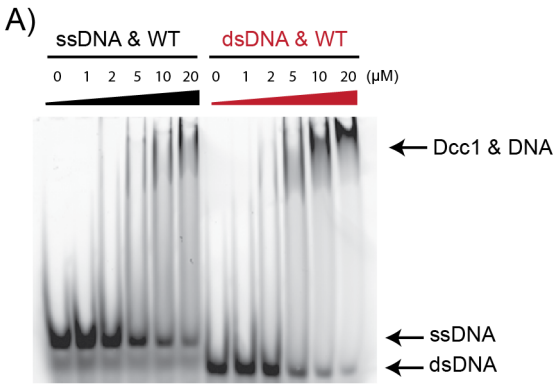
To further analyse if the WH domains of Dcc1 were able to interact with DNA an electrophoretic mobility shift assay (EMSA) was conducted (Figure 5.2A). Both ssDNA and dsDNA were kept at a constant concentration throughout the experiments with a fluorescently labelled tag located at the 5' end of each substrate, and for dsDNA the tag was located on only one strand (see methods). The Dcc1 C-terminal domain was titrated in at increasing concentrations, and samples were separated by native-PAGE. If Dcc1 bound to DNA a gel shift would be observed as Dcc1 would retard the DNA substrates in the gel.

The results from the EMSA, surprisingly, indicated that the C-terminal domain of Dcc1 could bind to both ssDNA and dsDNA (Figure 5.2A). This was shown as both substrates were retarded in the gel by increasing concentrations of Dcc1. Furthermore, Dcc1 had a higher affinity for dsDNA compared to ssDNA, as more dsDNA was retarded in the gel. The smeary bands present in the native gel meant it was not possible to determine binding affinities for Dcc1 to ssDNA or dsDNA.

To further clarify the interaction between Dcc1 and DNA five basic residues were identified, which were K326, K357, K364, K367 and R380 (Figure 5.2B). These residues were mutated to alanine and expressed and purified, along with a construct that contained a deletion of WH3 ($Dcc1^{\Delta WH3}$; residue 1-317) and full length Ctf8. The above mutated proteins were used in an EMSA to determine if they disrupted binding to DNA (Figure 5.3). As expected, $Dcc1^{\Delta WH3}$ completely abolished both ssDNA and dsDNA binding. In addition, mutations in K364A, R367A and R380A showed a mild reduction in both ssDNA and dsDNA binding (Figure 5.3), which suggests the binding site for ssDNA and dsDNA must overlap. To increase the effect of individual point mutation the triple mutation K364A, R367A and R380A was expressed and purified and the EMSA repeated. Together these mutation completely abolished DNA binding, which was equivalent to the $Dcc1^{\Delta WH3}$ deletion (Figure 5.3).

On further analysis of the $Dcc1^{\Delta WH3}$ construct a small gel shift was observed with ssDNA but not dsDNA when a lower concentration of DNA was used and the gel was overexposed (Figure 5.4A and 5.4B). This suggests that WH1 and WH2 contribute

to ssDNA binding, which is enhanced in the presence of WH3. Unfortunately, a mutation in WH2/WH1 that reduced the affinity of Dcc1 to ssDNA could not be identified.



B)

```
S.cerevisiae RLRIPFIAQWYGIQALRKYVSGISMPIDEFLIKWKSLFPPFFPCDIDIDMLRGYHF---K 307
S.pombe      NIDLELTTFWYGRFLLEG---INSISSDEFIQLWDNRLPYPCKGLPSLNLLKGYYF---H 281
X.laevis     ALDEDKICRATALMLLQN---AVKFNLAEFQEVWQQSVPDGM--NTRLDQLKGLALVDRT 313
M.musculus   ELDADKICRVTAEMLLQN---AVKFNLAEFQEVWQQSVPEGM--TTRLDQLKGLALVDRN 322
H.sapiens    ELDADKICRAAARMLLQN---AVKFNLAEFQEVWQQSVPEGM--VTSLDQLKGLALVDRH 316
              :         .   *.      .:   **   *.. .*        :: *:*  :

S.cerevisiae PTDKTVQYIAKSTLPMDPKERFKVLFRLQSQWDLEDIKPLIEELNSRGMKIDSFIMKYAR 367
S.pombe      DTPNTIQYLSGDQLPREPSRRFQSLFQLKSKWLYEELWSFVKDLALSKSRVEALILKYGR 341
X.laevis     SRPETIFLLQTEDLPEDTQERFNTLFGMREKWTEADIAPYIQDLCGEKQTIGALLTKYAR 373
M.musculus   SRPEIIFLLKVDDLPEGTQDRFNSLFSLREKWTEEDITPYIQDLCGEKQTIGALLTKYSR 382
H.sapiens    SRPEIIFLLKVDDLPEDNQERFNSLFSLREKWTEEDIAPYIQDLCGEKQTIGALLTKYSH 376
                : :  :  . **   . **: ** ::.:*   ::   :::*      : ::: **.:

S.cerevisiae RKRLGKKTVVTSR---- 380
S.pombe      KQTSRDGVYINSRGTW- 357
X.laevis     SSMQNGIKLFNSRRPLS 390
M.musculus   SSMQNGIKVYNSRRLIS 399
H.sapiens    SSMQNGVKVYNSRRPIS 393
              .        .**
```

Figure 5.2. DNA Binding of Dcc1.

A) EMSA analysis of Dcc1 with both ssDNA and dsDNA. Dcc1 was titrated at increasing concentrations into a constant amount of either ssDNA or dsDNA. Both substrates were fluorescently labelled and the concentration of Dcc1 used in each sample is indicated above the gel. B) Sequence alignment of the C-terminus of Dcc1 from yeast to human arrows indicate the basic residues mutated.

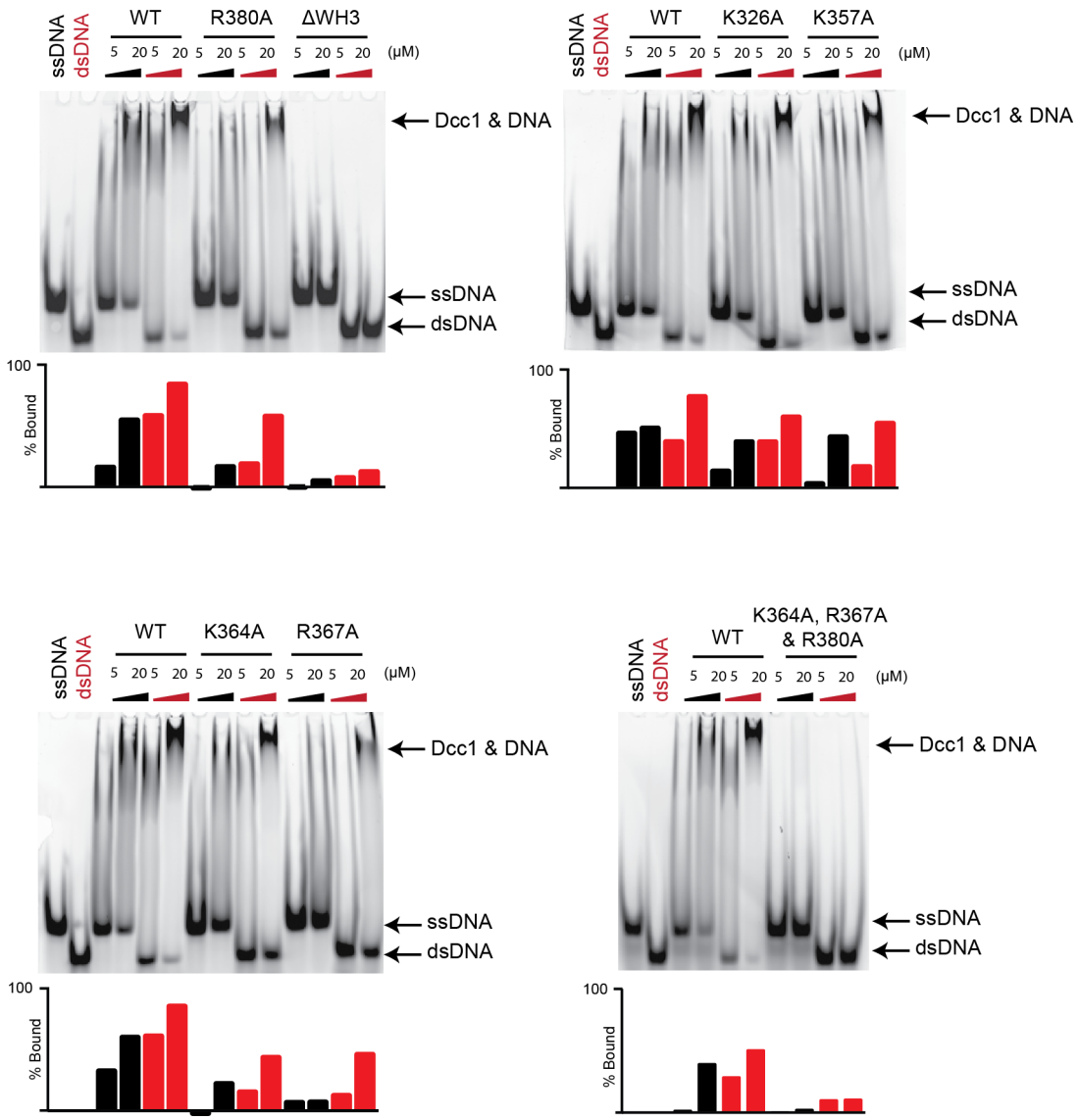


Figure 5.3. DNA Binding of Mutant Dcc1.

Mutant proteins were analysed against wild type protein for their effect on both ssDNA and dsDNA binding. The mutants used are indicated above the gel along with their concentrations. The bar charts below each gel were determined by analysing the amount of DNA that was not retarded compared to the control ssDNA and dsDNA.

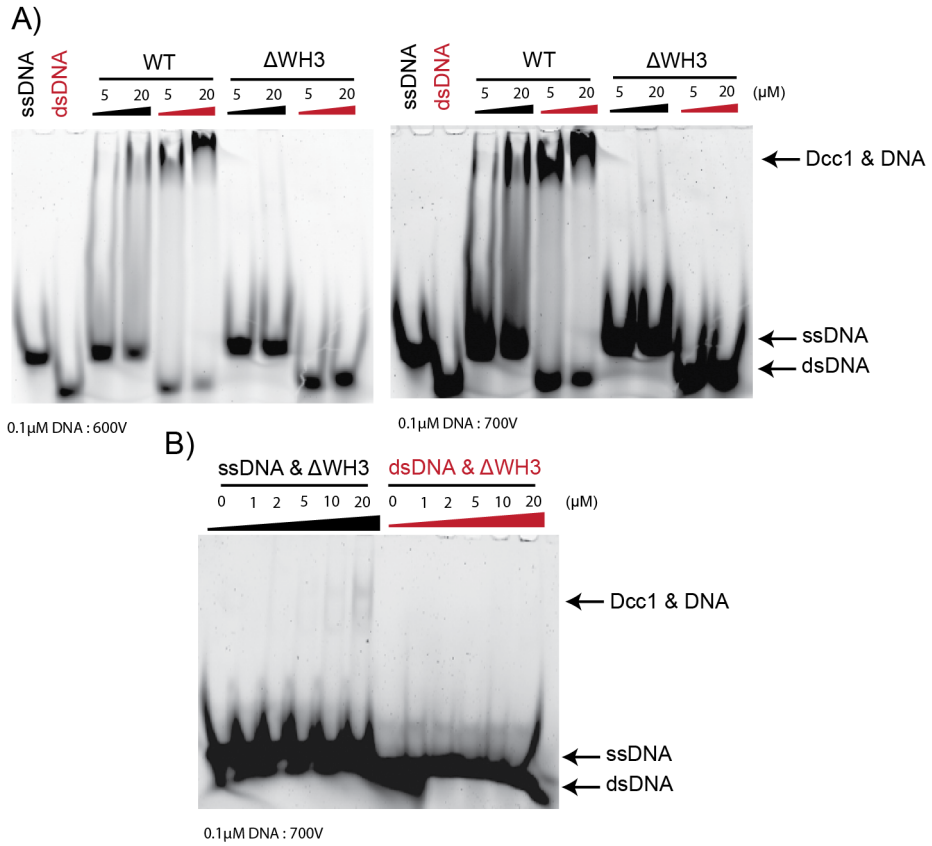


Figure 5.4. ssDNA Binding of WH1 and WH2 of Dcc1.

A) Analysis of WT and Dcc1$^{\Delta WH3}$ viewed at different voltages. The concentration of Dcc1 is labelled above the gel. B) Dcc1$^{\Delta WH3}$ was titrated at increasing concentrations against a constant concentration of ssDNA and dsDNA. The concentration of Dcc1$^{\Delta WH3}$ is indicated above the gel.

5.3 $Ctf18^{C}$-Dcc1-Ctf8 can Interact With Pol2

Previous studies identified that RFC^{Ctf18} was able to physically interact with polymerase-ε via pull down experiments, and this was shown to be conserved from yeast to humans (García-Rodríguez et al., 2015; Murakami et al., 2010; Okimoto et al., 2016). Furthermore, analysis using co-immunoprecipitation experiments revealed that the heterotrimer formed from Ctf18, Dcc1 and Ctf8 was required for the interaction with the N-terminal region of Pol2, which is the catalytic subunit of polymerase-ε (García-Rodríguez et al., 2015).

From the above information and the structure of $Ctf18^{C}$-Dcc1-Ctf8 solved in this thesis it is likely that Pol2 binds to the 'triple' β-barrel formed from Dcc1, Ctf8 and Ctf18. To test if this is the case, the *in vitro* interaction between the N-terminal region of Pol2 (residue 1-544) and $Ctf18^{C}$-Dcc1-Ctf8 was analysed with the use of a glycerol gradient (Figure 5.5). As predicted, the results indicated that the N-terminal region of Pol2 could interact with the heterotrimer (Figure 5.5A and B).

To further test if Ctf18 is required for the interaction with Pol2 the experiment was repeated with only the Dcc1-Ctf8 sub-complex (Figure 5.5B). The results revealed that the Dcc1-Ctf8 sub-complex could not interact with Pol2, which indicated the importance of the C-terminus of Ctf18 in binding Pol2.

To determine if the three WH domains of Dcc1 contribute to the interaction with Pol2, the glycerol gradients were repeated with $Ctf18^{C}$-$Dcc1^{\Delta 3}$-Ctf8 and $Ctf18^{C}$-$Dcc1^{\Delta 2\text{-}3}$-Ctf8 (Figure 5.5B). Unfortunately, $Ctf18^{C}$-$Dcc1^{\Delta 1\text{-}3}$-Ctf8 could not be expressed in *E. coli* and was not tested. Unexpectedly, the results revealed that both $Ctf18^{C}$-$Dcc1^{\Delta 3}$-Ctf8 and $Ctf18^{C}$-$Dcc1^{\Delta 2\text{-}3}$-Ctf8 could not interact with the N-terminal region of Pol2, which suggested that WH3 of Dcc1 must also aid in binding Pol2. The above results implied that the N-terminus of Pol2 must bind across the trimerisation domain formed from $Ctf18^{C}$, Dcc1 and Ctf8.

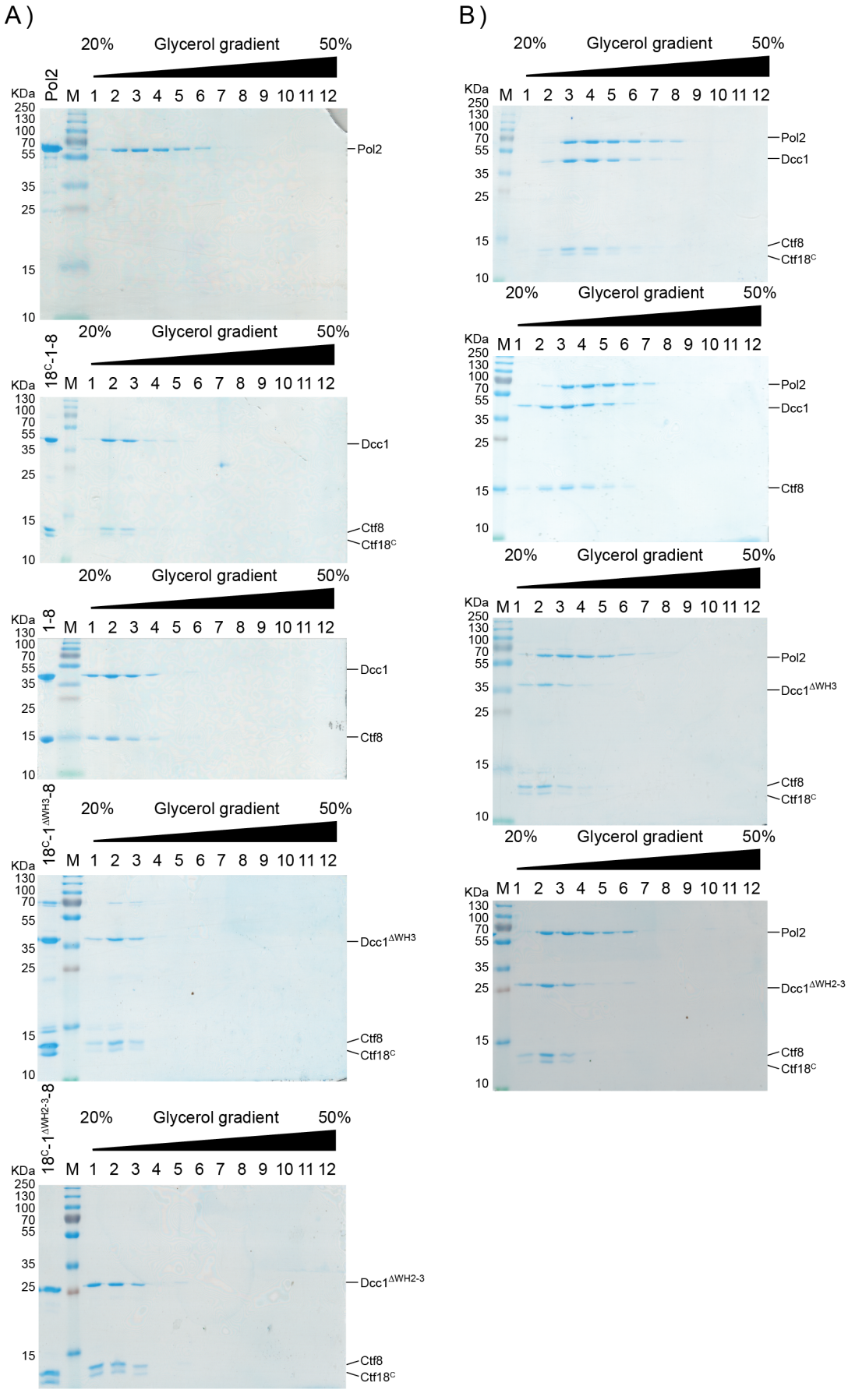

A)
B)
20%
Glycerol gradient
50%
KDa
M 1 2 3 4 5 6 7 8 9 10 11 12
Pol2
250
130
100
70
55
35
25
15
10
Pol2
Dcc1
Ctf8
Ctf18C
18C-1-8
1-8
18C-1ΔWH3-8
18C-1ΔWH2-3-8
Dcc1ΔWH3
Dcc1ΔWH2-3

Figure 5.5. Glycerol Gradients Between Pol2 and Ctf18^{C}-Dcc1-Ctf8
A) The SDS-PAGE analysis of the glycerol gradients of the N-terminus of Pol2, Ctf18^{C}-Dcc1-Ctf8 (18^{C}-1-8) and Dcc1-Ctf8 (1-8), Ctf18^{C}-Dcc1$^{\Delta 3}$-Ctf8 (18^{C}-1 $^{\Delta 3}$-8) and Ctf18^{C}-Dcc1$^{\Delta 2-3}$-Ctf8 (18^{C}-1 $^{\Delta 2-3}$-8). B) The SDS-PAGE analysis of the glycerol gradients of Dcc1 constructs with Pol2. The fraction number is presented above the gel.

5.4 Dcc1 is Involved in the Recruitment of Ctf18 to Chromatin

Work in this section has been conducted by both Dr. Catarina Samora and Hon Liu (see footnotes).

Previous work indicated that cells have a delayed activation of S-phase checkpoints in hydroxyurea (HU) when Ctf18, Dcc1 or Ctf8 were deleted (Crabbé et al., 2010). The DRC is stimulated by an accumulation of ssDNA; therefore, I hypothesised that the WH domains of Dcc1 could be responsible for recruiting RFC^{Ctf18} to sites of replication stress. To test this, ChIP-qPCR was conducted with an anti HA-antibody against HA-Ctf18 [1] (the ChIP-qPCR was conducted by Dr. Catarina Samora; therefore, this method is not presented in this thesis). The amount of Ctf18 was measured at three early and one late firing replication origins. The late firing origins should be inhibited in HU and have a much lower amount of Ctf18 compared to the early firing origins. The results indicated that *dcc1*Δ cells had a significantly reduced level of Ctf18 at the early firing origins (Figure 5.6A). This result was also the same for Dcc1$^{\Delta WH3}$; although, this mutant did not fully replicate the *dcc1*Δ. This is unsurprising, as the Ctf18-Dcc1-Ctf8 heterotrimer interacts with polymerase-ε and could be recruited to stalled replication forks by multiple mechanisms. To check that Dcc1$^{\Delta WH3}$ was still stable the gel filtration profiles of both Ctf8-Dcc1 and Ctf8-Dcc1$^{\Delta WH3}$ were analysed (Figure 5.6B). The results indicated that Ctf8-Dcc1$^{\Delta WH3}$ was soluble, and as expected, its elution volume is slightly later than Dcc1-Ctf8 because Ctf8- Dcc1$^{\Delta WH3}$ has a lower molecular weight.

To further analyse the effect of a deletion of WH3 on the DRC the sensitivity of *dcc1*Δ, Dcc1$^{\Delta WH3}$, Dcc1$^{\Delta WH2-3}$ and Dcc1$^{\Delta WH1-3}$ to HU was compared to WT cells[2] (Figure

1. ChIP-qPCR was carried out by Dr. Catarina Samora in Dr. Frank Uhlmann's group.
2. The sensitivity to HU, and the Rad53 phosphorylation assay was carried out by Hon Liu in Dr. Frank Uhlmann's group.

5.7A). As expected all the deletions caused an increase in sensitivity to HU. Additional analysis was conducted by measuring the rate of Rad53 phosphorylation in HU, which is a method to measure the activity of the DRC (Figure 5.7B). The results indicated that all deletions caused a delay in Rad53 phosphorylation, which was similar to the *dcc1Δ* strain. This indicated an impaired checkpoint response.

Taken together these results suggest that WH3 of Dcc1 has a role in recruiting RFC^{Ctf18} to replication forks, either through binding directly to DNA or through its interaction with Pol2. In addition, all the WH domains of Dcc1 participate in checkpoint activation.

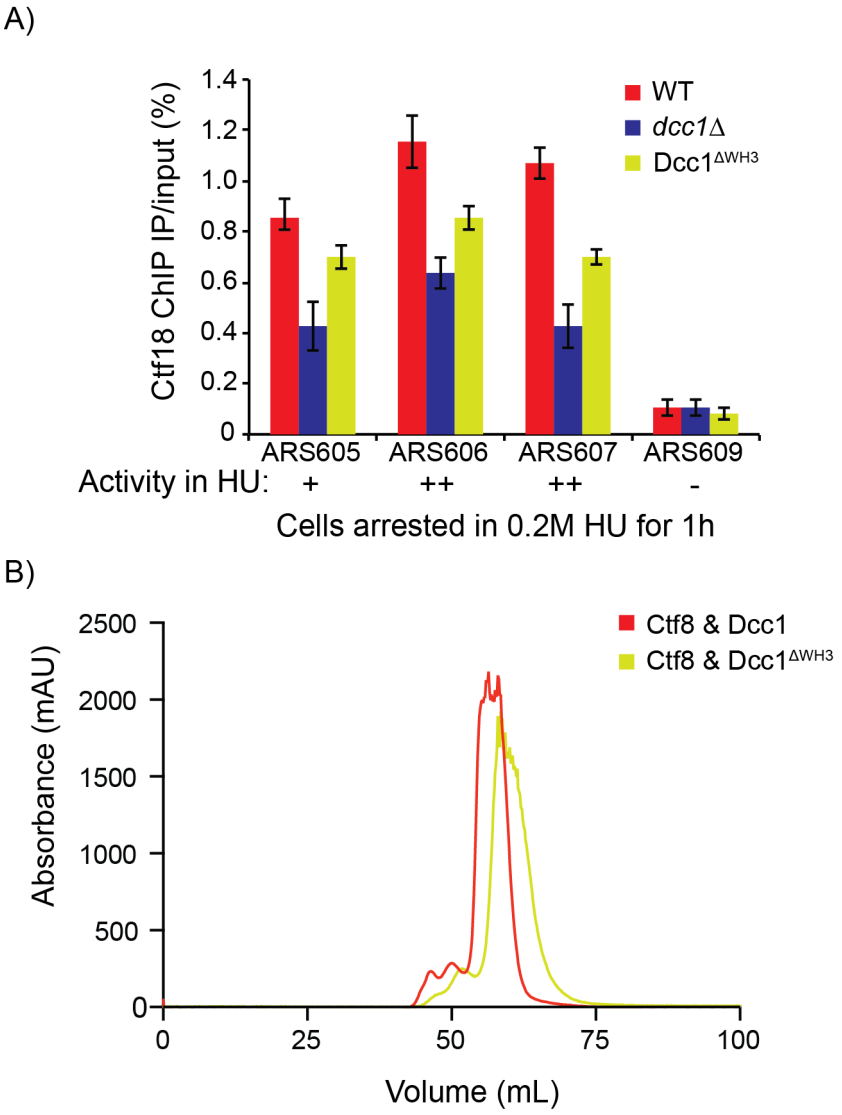


Figure 5.6. WH3 is Responsible for Localising RFC^{Ctf8} to Replication Origins.

A) The amount of Ctf18 present at replication origins was measured in HU-arrested cells. Three early firing origins (ARS605, ARS606 and ARS 607) were studied along with one late firing origin (ARS609). The analysis of Ctf18 was carried out by ChIP-qPCR with the use of an anti-HA antibody against HA-Ctf18. Three independent experiments were conducted and the means and standard errors plotted on the graph. B) Gel filtration profiles on a HiLoad 16/60 Superdex 75 column (GE healthcare) in 10 mM Tris 7.5, 150 mM NaCl and 0.5 mM TCEP of the two sub-complexes: Ctf8-Dcc1 and Ctf8-Dcc1 $^{\Delta WH3}$.

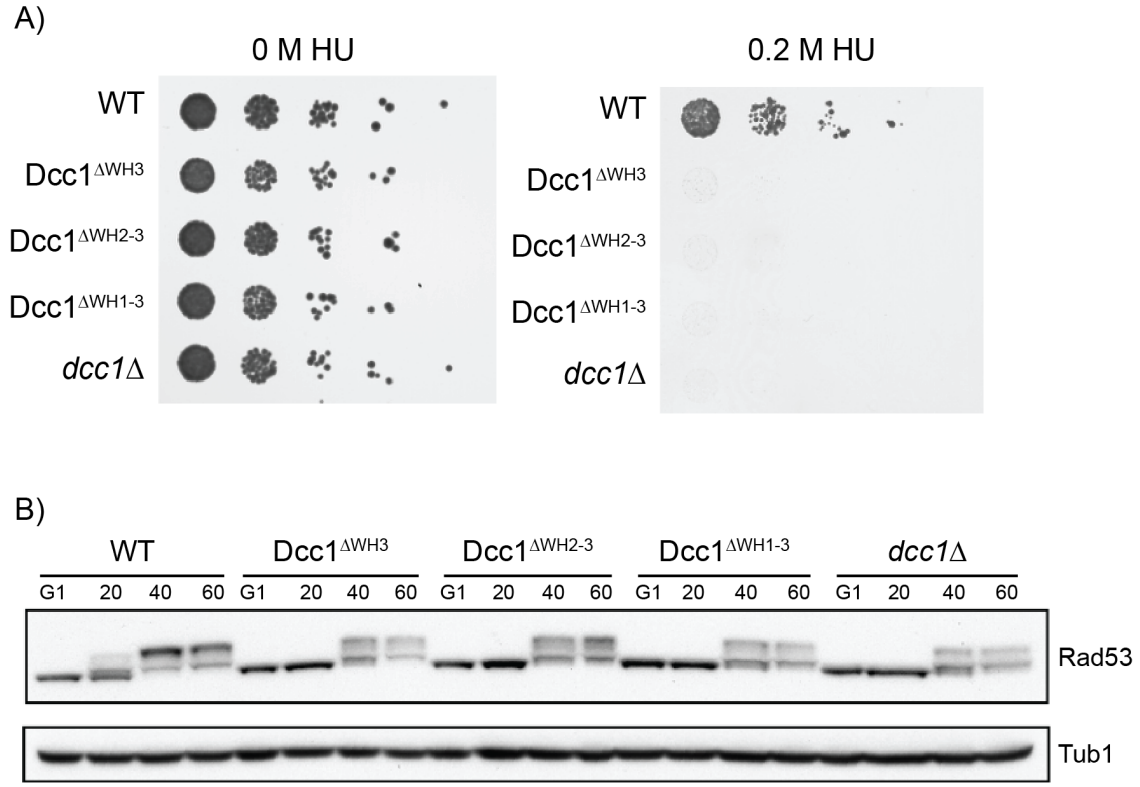


Figure 5.7. Analysis on the Function of the WH Domain of Dcc1 *in vivo*.

A) Spot assay in different hydroxyurea (HU) concentrations, at a 10-fold serial dilution. The cell strains and HU concentration used are indicated in the panel. B) Rad53 phosphorylation was monitored in the cell lines indicated above the gel. In G1, cells were arrested and then released in 0.2 M HU. Samples were taken at the indicated time point.

Chapter 6. Structure of xEco2 ACT Domain

6.1 Summary on the Function of Eco1

Sister-chromatid cohesion is known to be established during S-phase. This occurs through the action of Eco1, which acetylates the cohesin subunit Smc3 at residue K112 and K113 (K105 and K106 in higher eukaryotes). In cells the acetylation of K113 is more important than K112 acetylation, with the latter being almost dispensable for sister-chromatid cohesion (Zhang et al., 2008). Acetylation of these tandem lysine residues antagonises the anti-establishment activity of Wapl and Pds5 (Ben-Shahar et al., 2008). Recent work has suggested that acetylation of the tandem lysine motif negatively regulates the DNA induced hydrolysis of ATP by the head domain of cohesin (Murayama and Uhlmann, 2015).

Eco1 has a C-terminal Gcn5-relate N-acetyltransferase (GNAT) domain, which is well conserved throughout evolution (Ivanov et al., 2002). The N-terminus of Eco1 is divergent in both length and composition; however, all homologues contain both an N-terminal C_2H_2 zinc finger (ZnF) motif and a PIP box. The ZnF enhances acetylation of cohesin *in vivo* but it is still unclear how the ZnF performs this function (Onn et al., 2009). The N-terminal PIP box helps recruit Eco1 to the replisome via its interaction with PCNA (Moldovan et al., 2006). Apart from being spatially regulated, Eco1 is also degraded after S-phase through its ubiquitination by the combined action of Cdk-1 and Cdc4 (Lyons and Morgan, 2011).

Almost all eukaryotic organisms contain at least one copy of the gene encoding for Eco1 with functional specificity given by the varying N-terminus. For instance, the human genome contains two copies of Eco1 called human Eco1 (hEsco1) and human Eco2 (hEsco2). hEsco2 is the somatic acetyltransferase; mutations in this gene cause the genetic diseases SC Phocomelia and Roberts Syndrome.

Not only can Eco1 acetylate the tandem lysine motif of Smc3 during S-phase, Eco1 can also target Scc1 in response to a DSB (Heidinger-Pauli et al., 2009) and Eco1 can auto-acetylate itself *in vitro* (Ivanov and Nasmyth, 2005). To improve our understanding on the specificity of Eco1 two *in vitro* acetylation assays were

developed. Firstly, the rate of K112 and K113 acetylation was monitored by MS, and secondly a substrate motif was generated from a list of Eco1 acetylated peptides identified from the *in vitro* acetylation of a range of cohesin subunits. To rationalise the results from the above two experiment the crystal structure of *X. laevis* Eco2 (xEco2) acetyltransferase (ACT) domain was solved with two conjugated substrate peptides. Finally, the role of acetylation was investigated, *in vitro*, with the use of acetylation mimics to determine the impact on DNA binding.

6.2 K112 Acetylation is Faster Than K113 Acetylation

Work in this section has been conducted in collaboration with the Protein Analysis and Proteomics Platform at the Francis Crick Institute (see footnotes).

The conserved tandem lysine residues of Smc3 become acetylated during S-phase. Previous genetic experiments discovered that K112 acetylation was less important than K113 acetylation (Zhang et al., 2008) as mutating K113R caused poor cell survival and increased chromosome mis-segregation, and K112R showed only a mild phenotype. To further understand the mechanism of Smc3 acetylation mass spectrometry (MS) was used in combination with an *in vitro* assay[3] (Figure 6.1A). The MS analysis observed the abundance of the K112/K113 peptide (TVGLKKDDYQLNDR). Acetylation of K113 would block trypsin cleavage between K112/K113; therefore, to simplify the analysis before trypsin cleavage in gel reductive di-methylation was performed (Boersema et al., 2009). This meant trypsin could no longer cleave between the two lysine residues even if K113 was not acetylated. In this analysis a methylated lysine (me2K112 or me2K113) indicates that this lysine residue is not acetylated.

The results from the MS analysis indicated that the abundance of me2K112me2K113 decreased throughout the time course (Figure 6.1B). Furthermore, both the mono-acetylated peptide AcK112me2K113 and the tandem acetylated peptide AcK112AcK113 increased throughout the experiment (Figure 6.1C and D). The mono-acetylated peptide me2K112AcK113 could not be identified in the MS analysis.

3. MS was carried out in collaboration with the Protein Analysis and Proteomics Platform at the Francis Crick Institute.

To further improve the analysis MS/MS was conducted along with parallel reaction monitoring (PRM) and Q-Exactive mass spectrometry. The y9 peptide (Figure 6.2A) produced through the breakage between K112 and K113 could be observed in the resulting spectrum. For instance, at 25.5 minutes the AcK112me2K113 (1194.57, y9) peptide was observed in the spectrum but the me2K112acK113 (1208.55, y9) peptide could not be identified (Figure 6.2B and C). The above results suggest that K112 acetylation occurs at a faster rate than the acetylation of K113.

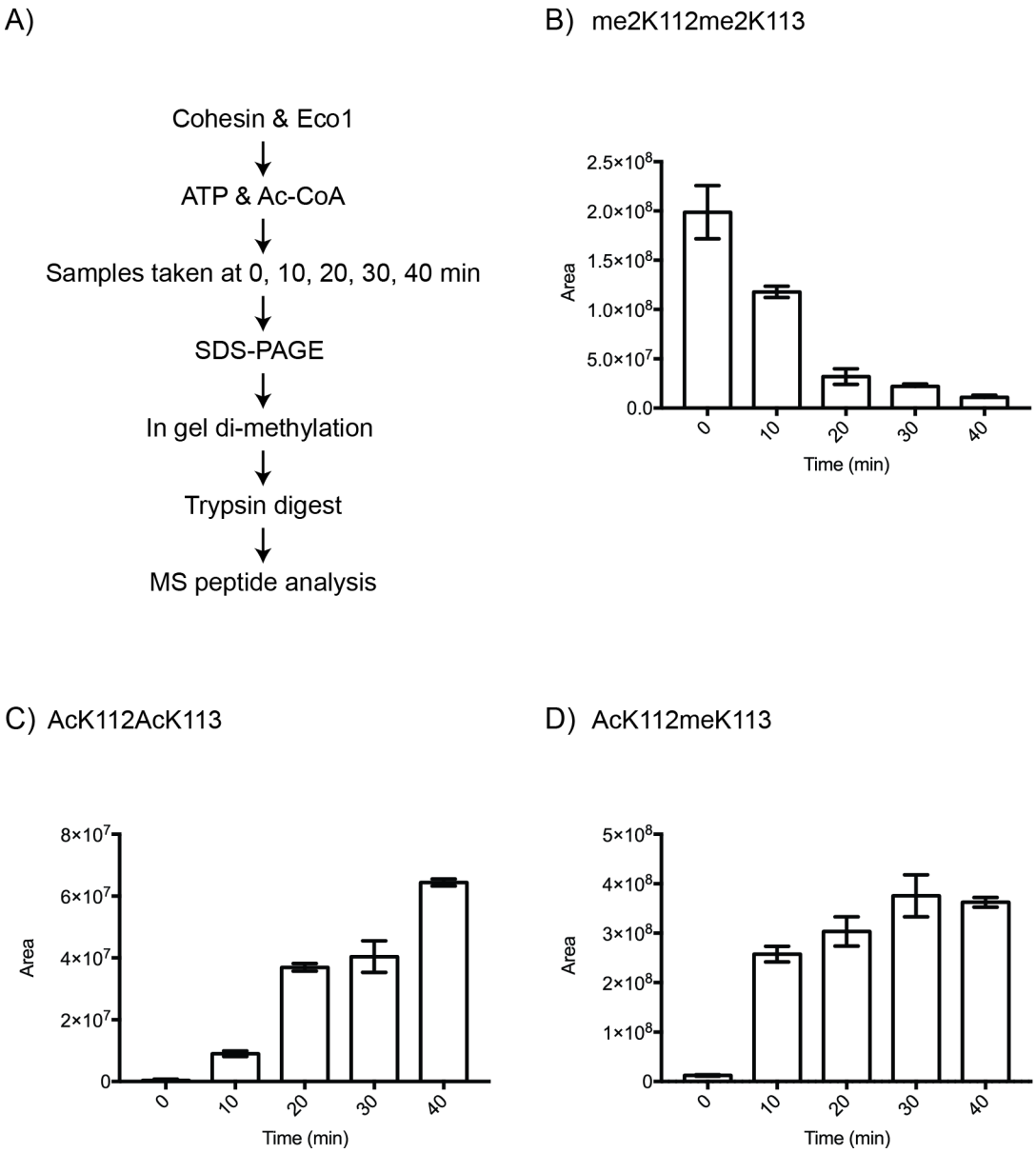


Figure 6.1. Analysis of the Acetylation of the Tandem Lysines of Smc3.

A) Description of the in vitro acetylation assay using di-methylation to improve the analysis. B) Abundance of no acetylation (me2K112me2K113) recorded at each time point. C) Abundance of K112 and K113 (AcK112AcK113) acetylation at each time point. D) Abundance of K112 (AcK112me2113) acetylation at each time point.

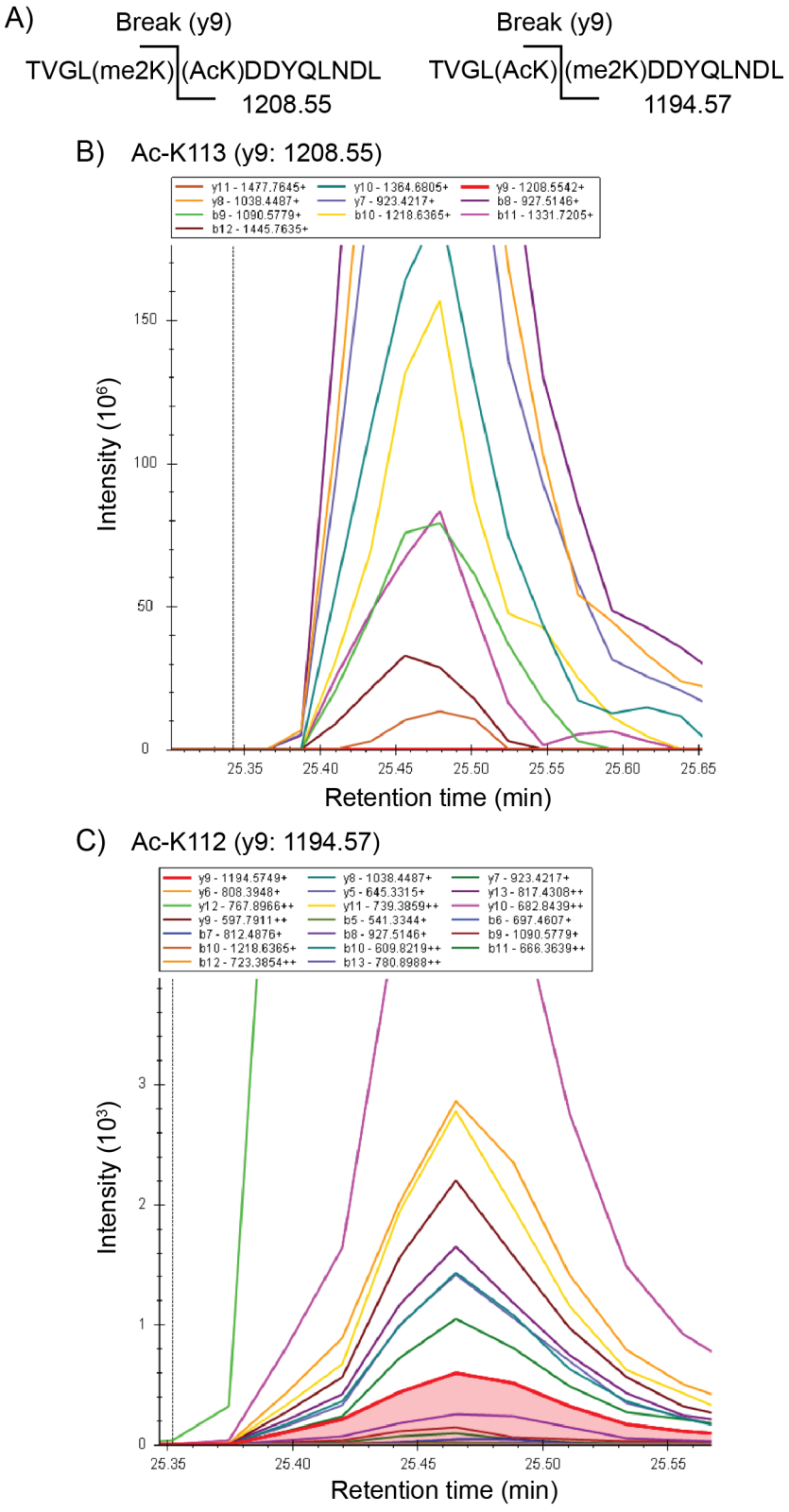


Figure 6.2. Abundance of Peptide Fragments.

A) The production of the y9 peptide in the MS/MS analysis through the breakage of the K112 and K113 peptide bond. B) Analysis of the abundance of different peptide ions,

AcK113 (1208.55) is shown in red C) Analysis of the abundance of different peptide ions, AcK112 (1194.57) is shown in red.

6.3 Eco1 can Acetylate a Variety of Substrates

Work in this section has been conducted in collaboration with the Protein Analysis and Proteomics Platform at the Francis Crick Institute (see footnotes).

Apart from acetylating the tandem lysine motif of Smc3, Eco1 is also able to both self-acetylate itself *in vitro* (Ivanov et al., 2002) and acetylate Scc1 in response to a DSB (Heidinger-Pauli et al., 2009). Neither Eco1 or Scc1 contain a conserved tandem lysine motif; therefore, it could be hypothesised that Eco1 has a broad substrate motif. To test this further, cohesin and its accessory subunits were recombinant expressed (Scc2-Scc4, Smc1-Smc3-Scc1-Scc3, Pds5 and Wapl) and acetylated *in vitro* by Eco1 *(*Figure 6.3A). The resulting acetylated proteins were analysed by MS to produce a list of lysine residues which became acetylated[4]. A control for each protein was used to determine acetylation which was not due to Eco1. For instance, acetylation of the proteins could have occurred during expression or purification.

The results from the MS analysis indicate that acetylated lysine residues were predominately flanked by an aliphatic (ϕ) and an acidic/polar (B/Z) residues. Furthermore, this motif occurred in either orientation suggesting Eco1 had a bi-direction motif. To further investigate this, the peptides were separated into three groups: forward, reverse and no-direction (Appendix 3). The iceLogo server (Colaert et al., 2009) was then used to determine if there was an abundance of a specific amino acid. The following amount of peptides were used in each class; 36 forward peptides (ϕ-AcK-B/Z), 42 reverse peptides (B/Z-AcK-ϕ) and 27 peptides with no identifiable direction (X-AcK-P/R) (Figure 6.3A, B and C). The mixed direction group were usually flanked by a proline or arginine which indicated the β-hairpin conformation seen in the Smc3-Scc1N structure (Gligoris et al., 2014) is an important aspect of Eco1 specificity.

4. MS was carried out in collaboration with the Protein Analysis and Proteomics Platform at the Francis Crick Institute.

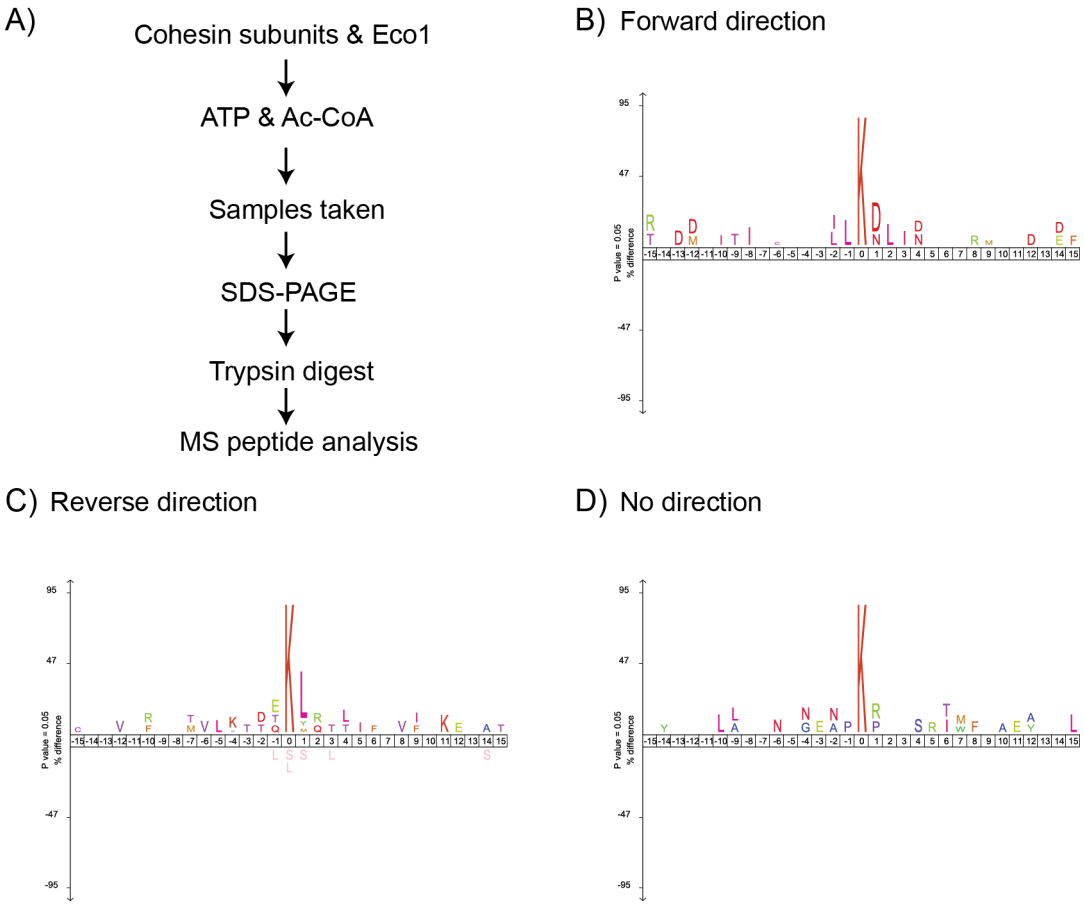


Figure 6.3. Generation of a Consensus Motif for Eco1 Acetylation.

A) Assay for analysis of in vitro acetylation of a range of cohesin subunits by Eco1. B) Analysis of peptides by the iceLogo server identified in the forward direction. C) Analysis of peptides by the iceLogo server identified in the reverse direction. D) Analysis of peptides by the iceLogo server identified with no direction.

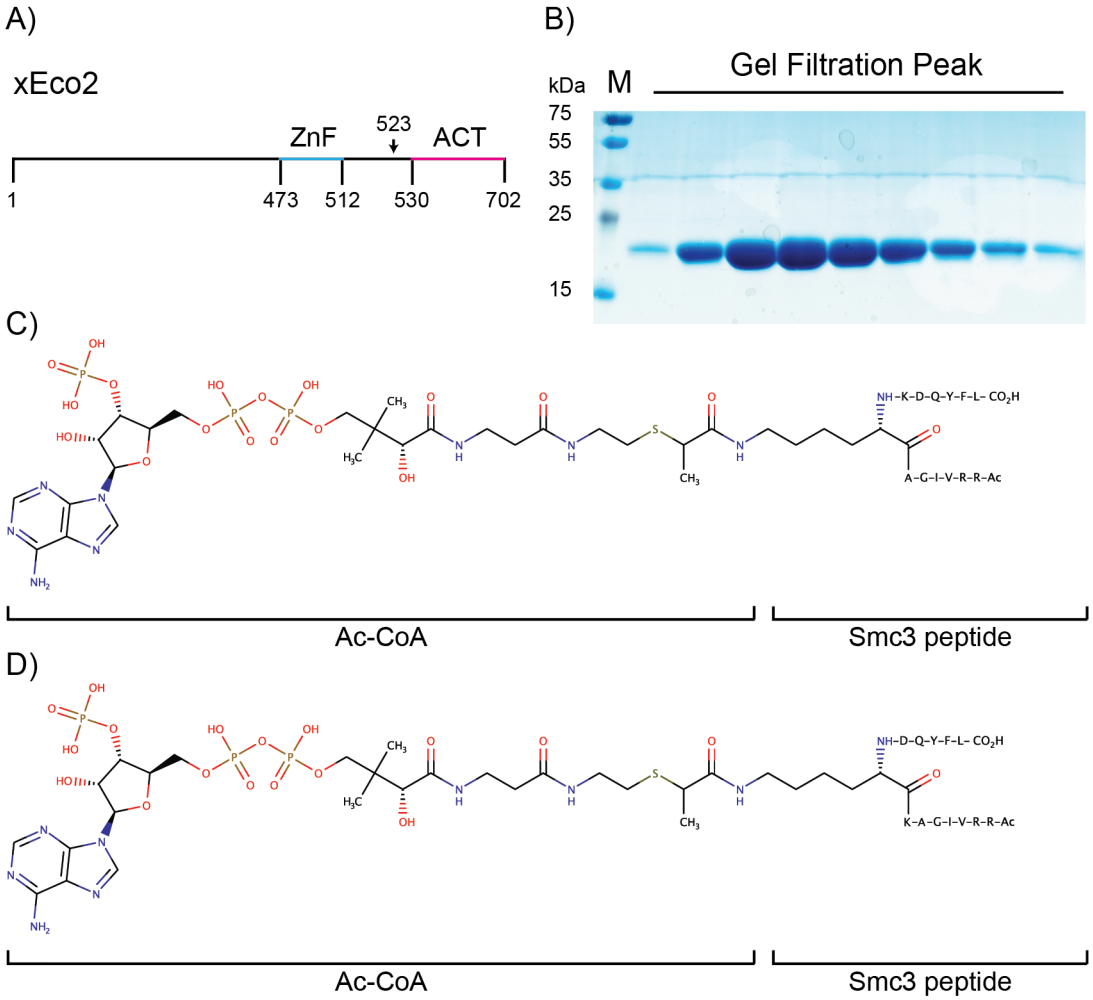
6.4 Crystallisation of xEco2 ACT Domain

Some work in this section has been conducted by Dr. Celine Bouchoux (see footnotes).

To gain further insights into Eco1 specificity the crystal structure of Eco1 bound to its canonical substrate was determined. To do this *X. laevis* Eco2 acetyltransferase (xEco2 ACT) domain was cloned (residue 523-702) based on the hEsco1 structure (Figure 6.4). This protein, xEco2, was chosen as the coding sequence for the gene was readily available in the laboratory. The protein expressed well and purified to a mono-species using a three step purification procedure (Figure 6.4B). The activity of this purified enzyme was not tested in this thesis.

The relative affinity of xEco2 to the head domain of cohesin was weak; therefore, to artificially increase the affinity two 13 residue peptides were synthesised with Ac-CoA conjugated to either K105 (Ac-CoA-K105) or K106 (Ac-CoA-K106) by an isopropionyl bridge (K112 and K113 in *S. cerevisiae*) (Lau et al., 2000; Poux et al., 2002) (Figure 6.4C and D). These peptides would mimic an intermediate binding state and help provide further evidence for the importance of the conserved tandem lysine motif (Figure 6.4E).

To crystallise xEco2 ACT domain, the protein was concentrated to 15 $mgmL^{-1}$. Initial crystal trials were conducted in a variety of commercially available screens with xEco2 alone or mixed with either conjugated peptide. Crystals of both xEco2 alone, Ac-CoA-K105 and Ac-CoA-K106 were obtained in variety of conditions (Figure 6.4F). xEco2 without peptide, Ac-CoA-K105 and Ac-CoA-K106 crystals diffracted to a maximum resolution of 3 Å, 2.4 Å and 1.9 Å respectively (Figure 6.4G). The three structures were solved by molecular replacement using either hEsco1 (Kouznetsova et al., 2016) or xEco2 ACT domain as a search model in Phaser (McCoy et al., 2007). An initial model was traced in AutoBuild (Adams et al., 2010) and multiple rounds of rebuilding and refinement were carried out in coot (Emsley and Cowtan, 2004) and Phenix.Refine (Adams et al., 2010) respectively. Full crystallographic statistics for data collection and refinement are given in Table 2.7, 2.8 and 2.9.



E)

```
S.cerevisiae LSRGDDEVTIRRTVGLKKDDYQLND 120
S.pombe      FPTGKSEVVLRRTIGLKKDEYSLDK 113
X.laevis     LPIDKEEVSLRRVIGAKKDQYFLDK 113
M.musculus   LPIDKEEVSLRRVIGAKKDQYFLDK 113
H.sapiens    LPIDKEEVSLRRVIGAKKDQYFLDK 113
             :   ..** :**.:* ***:* *:.
```

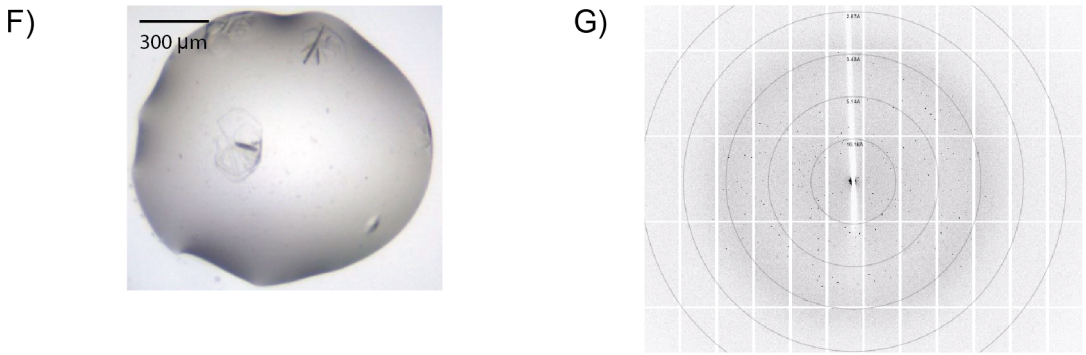


Figure 6.4. xEco2 ACT Expression and Crystallisation.

A) Secondary structure of xEco2. The ZnF and ACT domains are shown in blue and pink respectively. B) SDS-PAGE analysis of the final purity of xEco2 ACT domain in the third purification step. C) Peptide Ac-CoA-K105. Ac-CoA is conjugated to K105 of Smc3. D) Peptide Ac-CoA-K106. Ac-CoA is conjugated to K106 of Smc3. E) Sequence alignment of Smc3 target motif from yeast to humans. F) Crystallisation of xEco2 ACT. G) Diffraction of xEco2 ACT crystals at Diamond Light Source on beamline I02.

6.5 Overall Structure of xEco2 ACT Domain

xEco2 ACT domain forms a globular structure resembling the GNAT family, which is made from three α-helixes and seven β-strands (Figure 6.5). xEco2 ACT domain forms a structure with two globular regions termed the N- and C-lobe. CoA binds between these two domains interacting predominantly with β5 of the N-lobe and α3 of the C-lobe. The N-lobe is formed from β1, β2, α1, α2, β3, β4 and β5 (Figure 6.5); between β4 and β5 a conserved Eco1 specific insertion forms an extended β-hairpin structure. This hairpin structure is also seen in the human structure of hEsco1. The C-lobe contains α3, β6 and β7 as well as a C-terminal extension (C-extension), which undergoes conformational rearrangements depending on the substrate bound (see below). The two peptides interact in different conformations with xEco2; however, both peptides bind predominantly to the β-hairpin and C-extension.

The overall electron density of both peptides is of good quality. However, the density for peptide Ac-CoA-K105 is significantly better than Ac-CoA-K106 as for Ac-CoA-K106 side chain density can be seen for most residues (Figure 6.6) but for Ac-CoA-K105 only strong density for the peptide backbone could be observed (Figure 6.6A).

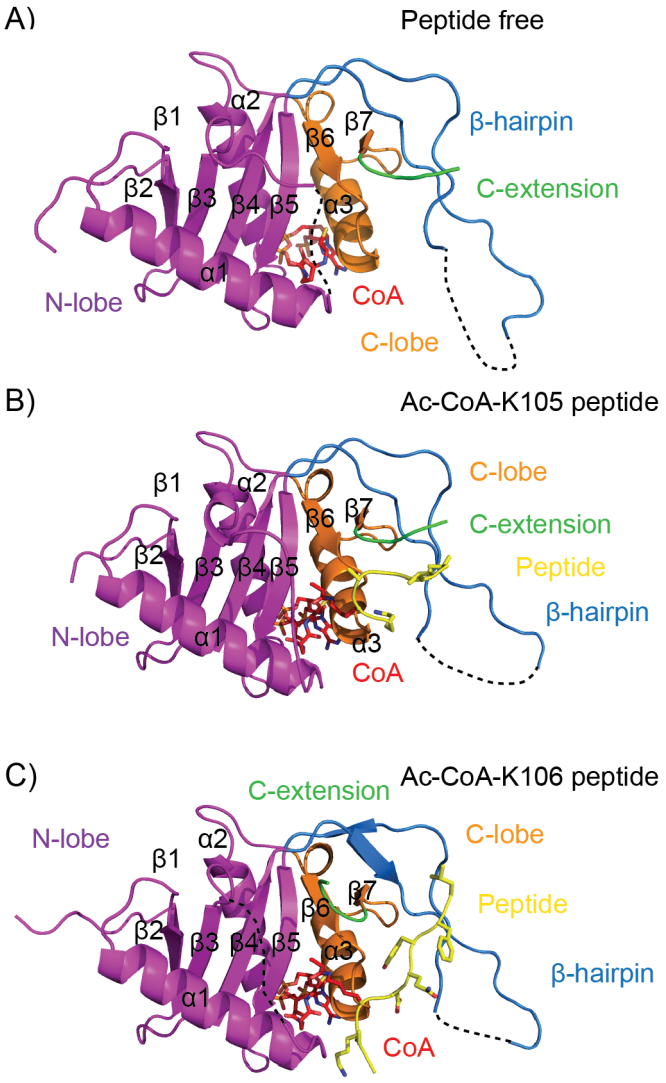


Figure 6.5. Overall Structure of xEco2 Bound to two Substrate Peptides.

A) The crystal structure of xEco2 ACT domain without peptide bound. The colouring is as follows; N-lobe (pink), C-lobe (orange), β-hairpin (blue), C-extension (green) and CoA (red). B) Crystal structure of xEco2-Ac-CoA-K105. The colouring is the same as in (A) and the peptide is shown in yellow. C) Crystal structure of xEco2-Ac-CoA-K106. The colouring is the same as in (B).

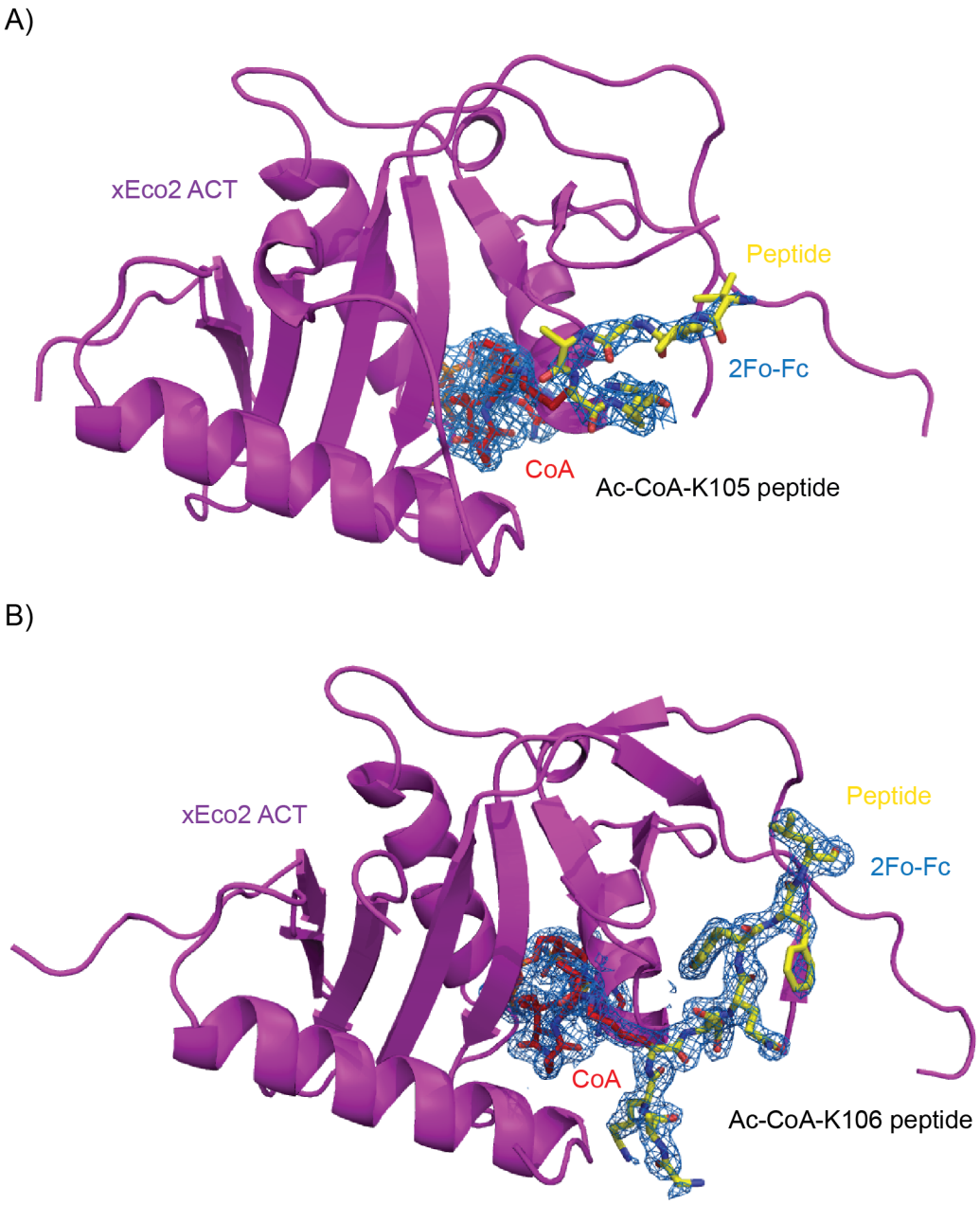


Figure 6.6. Peptide Density for Both K105 and K106 Crystal Structures.

A) Crystal structure of xEco2-Ac-CoA-K105 showing the electron density of a 2Fo-FC map of Ac-CoA-K105.The colouring is as follows; xEco2 ACT (pink) peptide (yellow), CoA (red) and 2Fo-Fc map (blue). B) Crystal structure of xEco2-Ac-CoA-K106 showing the electron density of a 2Fo-FC map of Ac-CoA-K106.The colouring is the same as in (A). Both maps are contoured to 1 σ.

6.6 xEco2 Uses a Conserved Glutamate as a Catalytic Residue

The authors of a previous study suggested that hEsco1 acetylates its targets using substrate assisted catalysis (Kouznetsova et al., 2016). They rationalised that a conserved aspartate on Smc3 acted as a general base for nucleophilic attack on the tandem lysine motif. This was supported by the observation that hEco1 was missing a conserved glutamate on β5 at residue 768, which is mutated to a conserved glycine in all Eco1 homologues. However, carrying out a structural alignment with xEco2 ACT and PCAF, a GNAT family member, shows that although the glutamate on β5 is missing the conserved E594 of xEco2 on β4 is in a related position. PCAF uses E570 as a general base for nucleophilic attack towards the target lysine; therefore, it could by concluded that E594 of xEco2 carries out the same function as this residue is only 0.1 Å further away from the substrate (Figure 6.7A). In addition, D107 of Smc3 in Ac-CoA-K106 forms crucial contacts with the β-hairpin indicating it is unlike to function in the catalytic mechanism. Taken together the above arguments suggests that hEsco1 and xEco2 do not operate by substrate assisted catalysis.

Previous work has also identified that hEsco1 ACT domain crystallises and behaves in solution as if it is a dimer (Kouznetsova et al., 2016). This dimer is formed from multiple contacts between both β-hairpins. However, xEco2 ACT domain crystallises with both conjugate peptides in a different conformation from hEsco1 raising the possibility that xEco2 does not form a dimer in solution. To test this, size exclusion chromatography and multi-angle laser light scattering (SEC-MALLS) was employed. The results indicated that xEco2 ACT domain behaves as a monomer in solution (Figure 6.7B) as xEco2 ACT domain had a calculated mass of 25.5 kDa, which is similar to its predicted mass of 21.8 kDa. The addition of Ac-CoA-K106 (2.5 kDa) to xEco2 increased the observed mass to 26.7 kDa suggesting the peptide did not change the oligomeric state of xEco2. The above results reveal that xEco2 ACT domain is a monomer in solution, which is different from hEsco1.

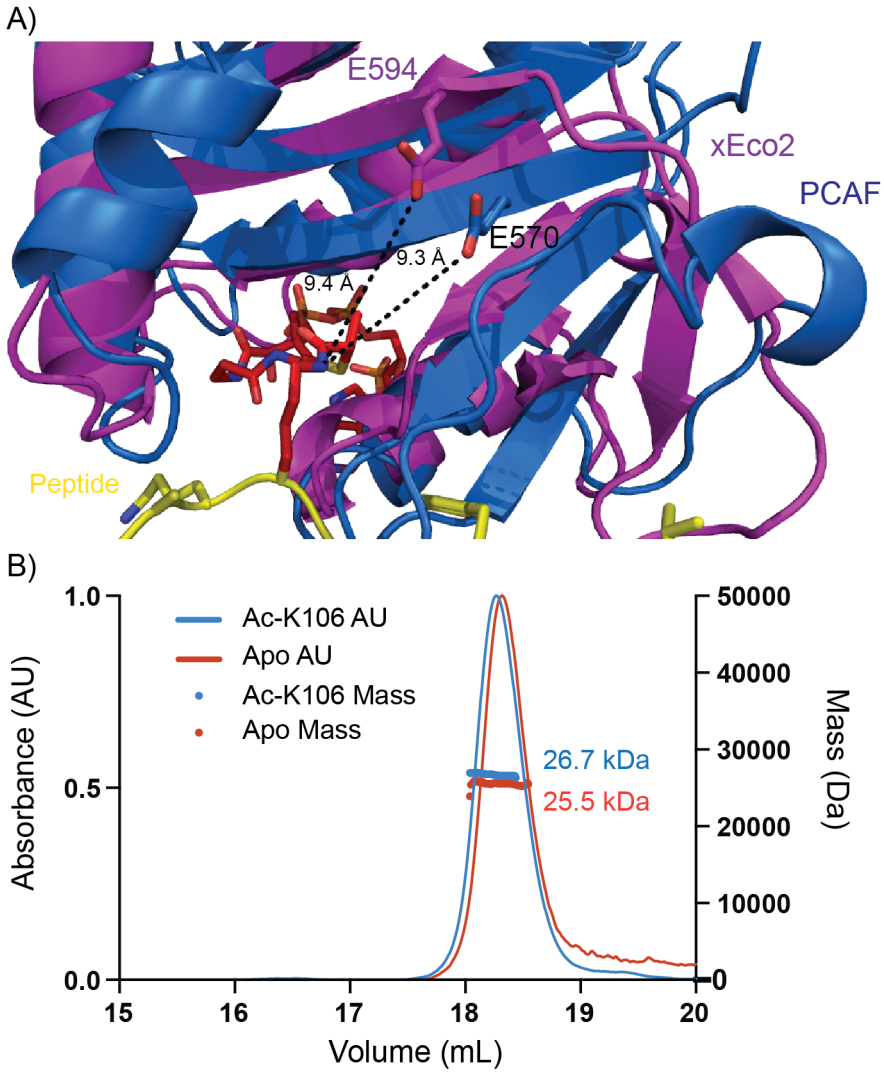


Figure 6.7. Structural Alignment of xEco2 with PCAF

A) Crystal structure of xEco2-Ac-CoA-K106 (pink) aligned with human PCAF (blue). E594 of xEco2 and E570 of PCAF are shown in stick representation. Ac-CoA-K106 peptide is shown in yellow and CoA is shown in red. Distances between the lysine and the glutamate residue are indicated in black. B) Analysis of the oligomeric state by SEC-MALLS of xEco2 ACT domain alone (red) and with Ac-CoA-K106 (blue). The masses of both peaks are indicated in kDa.

6.7 xEco2 Binds to the two Peptides in Different Conformations

All Eco1 homologues contain a range of conserved structural features, which are important for the interaction with Smc3 (Figure 6.8) such as the β-hairpin, C-extension and a hydrophobic pocket. The two peptides use different binding mechanisms to xEco2 ACT domain which, suggests their relative kinetics could be distinct. In addition, the Ac-CoA-K105 structure has weaker electron density than Ac-CoA-106 (Figure 6.6A and B), which implies that Ac-CoA-K105 has a weaker affinity to xEco2 than Ac-CoA-K106. Ac-CoA-K105 forms a β-hairpin structure, with position -1 (P-1) to P-4 from the conjugated lysine forming an anti-parallel β-sheet with the C-extension of xEco2 (Figure 6.9A). The unlinked lysine at P+1 in this structure interacts with D677 and W623 via a salt bridge and hydrogen bond respectively. The salt bridge interaction between D677 and K106 suggests the importance of having a basic residue at P+1 in the peptide motif identified with no direction (Figure 6.3D). Furthermore, a hydrophobic interaction was observed between I102 and V702 from the peptide (P-3) and xEco2 respectively. Further interactions were seen between L562 and F564 of xEco2 and the conjugated K105, which also form the base of the CoA binding pocket.

The Ac-CoA-K106 peptide forms an extended β-conformation when it is bound to xEco2 ACT (Figure 6.9B), which is similar to how the corresponding residues look in the Smc3-Scc1N crystal structure (Gligoris et al., 2014). The relative position of xEco2 on the Smc3 peptide has changed by 180 ° when comparing the two peptides. Ac-CoA-K106 binds to xEco2 ACT through multiple interaction with the second strand of the β-hairpin. In addition, key interactions were found between the two loops at the C-terminus of α1 and between the loop connecting α3 and β6. Like the other peptide structure L562 and F564 of xEco2 interact with CoA at the base of its binding pocket. However, these residues also bind through their aliphatic moieties to K105 and A104 at P-1 and P-2 respectively, from the conjugated lysine (K106). The residue D107 of the Smc3 peptide at P+1 interacts with xEco2 via a hydrogen bonds with the indole side chain of W623. In addition, D107 also forms a salt bridge with R621 of xEco2. Taken together these results support the peptide motif analysis,

which shows that Eco1 required P+1 to be acidic or polar, and both P-1 and P-2 to be aliphatic (Figure 6.3B and C).

The strong interaction between xEco2 and Ac-CoA-K106 is further supported by the binding of Y109 from the Smc3 peptide to a hydrophobic pocked formed from V602 of the C-extension; W623 and C625 from the β-hairpin; and V697 from β7 (Figure 6.9B). Destabilising this hydrophobic pocket by mutating both W192A and M193A in yeast, which correspond to W623 and C625 respectively in xEco2 causes a reduction in viability of yeast cells and an abolishment of Smc3 acetylation[5] (Figure 6.9C and D). Deleting the β-hairpin (yeast 189-198) also showed the same phenotype, which indicates the importance of both structural features (Figure 6.9C and D).

6.7.1 Roberts syndrome and yeast temperature sensitive mutations

Previous work has identified that the yeast temperature sensitive mutation G211D caused an increase in chromosome mis-segregation (xEco2; G635). Mutating G635 on xEco2 would disrupt CoA binding reducing the activity of this enzyme. This is supported by the reduction in viability and loss of the acetylation of Smc3 (Figure 6.9C and D) when the mutation G211D on Eco1 was introduced to yeast cells. In addition, the residue corresponding to the Roberts syndrome mutation in xEco2 (xEco2 W640; hEsco1 W639G; Eco1 W216G) would disrupt the integrity of the ACT domain. This was reinforced by the results in yeast cells, which show mutating W216G causes yeast cells to have a lower viability and a reduction in the acetylation of Smc3 (Figure 6.9C and D).

6.7.2 Rearrangement of C-extension upon substrate binding

When xEco2 is bound to Ac-CoA-K105, the peptide forms a β conformation with the C-extension of xEco2 (Figure 6.10A). This means the C-extension is in the same conformation for both the peptide free and Ac-CoA-K105 structures (Figure 6.10). However, upon binding Ac-CoA-K106 the C-extension is displaced and flips almost 180 ° (Figure 6.10B and C). This extreme movement is caused by the displacement

5. Yeast *in vivo* work (Figure 6.9C and D) was carried out by Dr. Celine Bouchoux in Dr. Frank Uhlmann's group.

of F700 by Y109 at P+2. The change in C-extension confirmation upon substrate binding is similar to the activation of Cdks, which undergo T-loop switching (Morgan, 1997).

The difference in substrate binding of K105 and K106 (K112 and K113 in yeast) to xEco2 ACT domain could explain the difference in the rates observed in the MS experiments. For instance, upon binding K106 the conserved phenylalanine has to be displaced by the tyrosine at P+2, which could be kinetically unfavourable. However, upon binding K105 the C-extension maintains the same conformation, possible making this substrate more favourable.

6.7.3 Docking of both xEco2 substrate structures onto Smc3

The peptide confirmation observed in Ac-CoA-K105 and Ac-CoA-K106 structures are similar to the corresponding regions in the Smc3-Scc1N structure (Gligoris et al., 2014). This further provides evidence that the two peptide structures observed in this thesis are likely to be correct. In addition, it permitted the docking of xEco2 ACT domain onto the yeast structure of Smc3 in two positions (Figure 6.11A and B). The structural analysis reveal that switching from K105 to K106 causes Eco1 to rotate as it translates along the β-hairpin of Smc3.

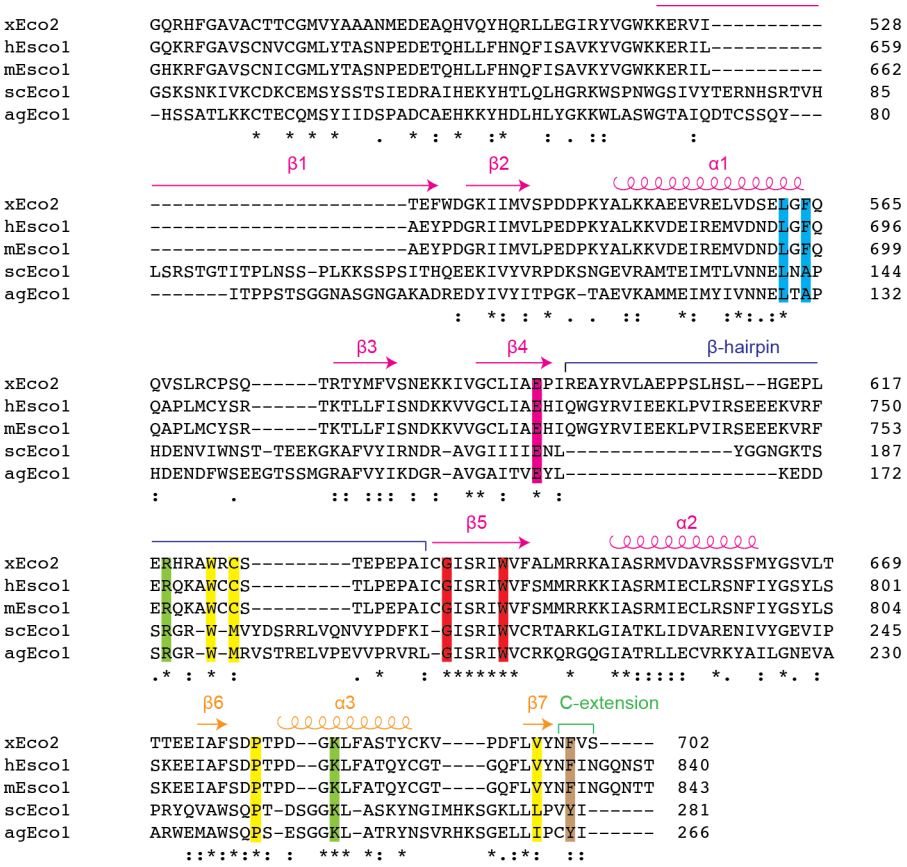


Figure 6.8. Sequence Alignment of Eco1 From Yeast to Humans.

Structural features are identified above the sequence alignment with the colouring as follows N-lobe (pink), β-hairpin (dark blue), C-lope (orange) and C-extension (green). Key conserved residues are indicated throughout the sequence alignment. In blue are residues important for binding CoA. In pink shows the conserved catalytic resides. In yellow are shown the residues that form the hydrophobic pocket, crucial for substrate binding. In red, the conserved glycine and tryptophan are shown, which are found mutated in a yeast temperature sensitive strain and Roberts syndrome respectively. In green, are additional basic residues important for substrate binding. In brown is the conserved aromatic residue found in the C-extension.

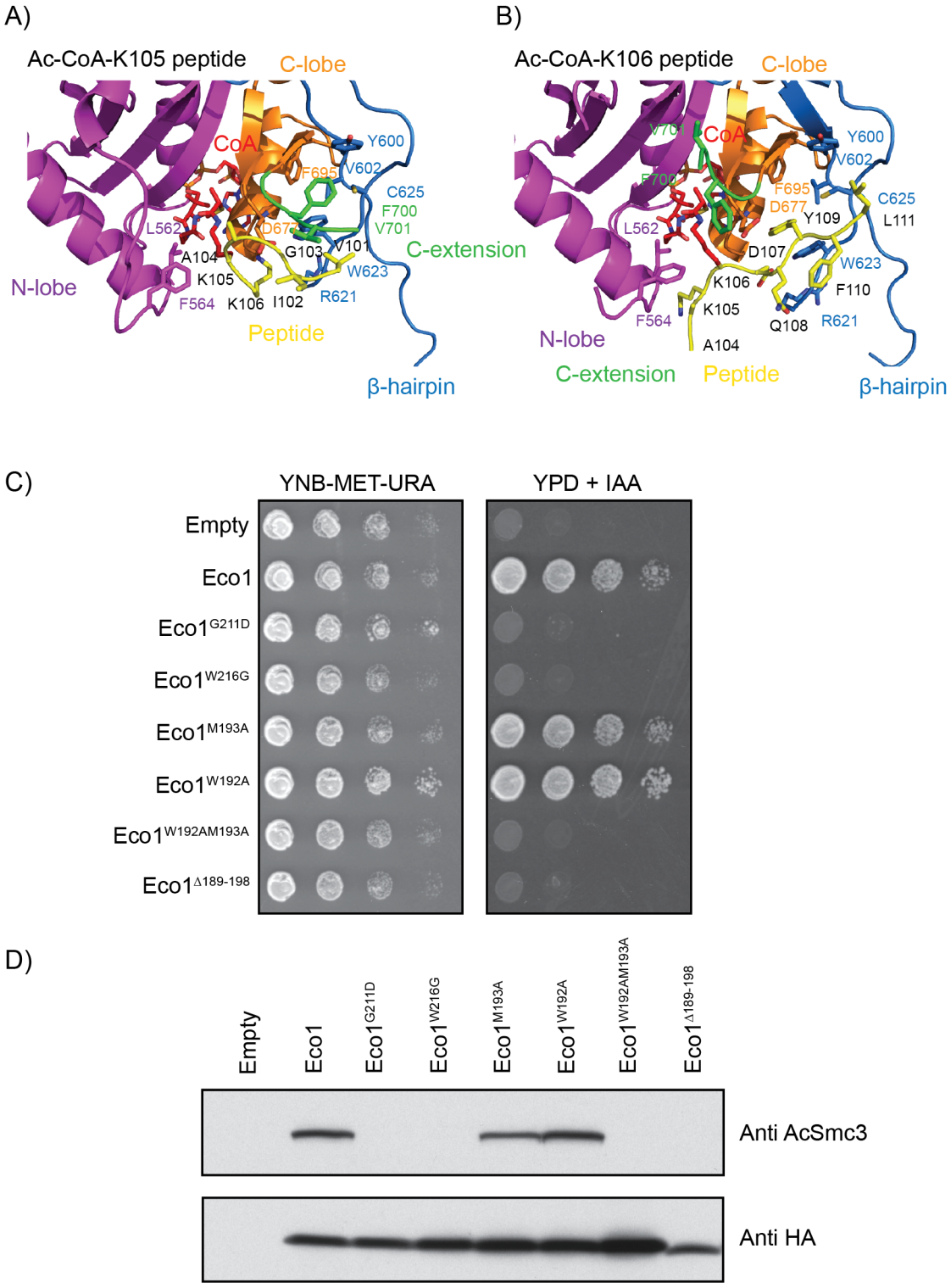


Figure 6.9. Difference in Binding Between the Peptides to xEco2.

A) Structure of xEco2 ACT domain bound to Ac-CoA-K105 with key residues shown in stick representation. The colouring is as follows N-Lobe (pink), C-lobe (orange), β-hairpin (blue), C-extension (green), CoA (red) and peptide (yellow). B) Structure of xEco2 ACT bound to Ac-CoA-K106 with key residues shown in stick representation. The colouring

is the same as in (A). C) Analysis of the survival of yeast in Eco1 mutations strains using a spot assay. Mutations are indicated in the panel. D) Analysis of acetylation of Smc3 by Eco1 mutations in vivo using a western blot. Mutations are indicated above the gel.

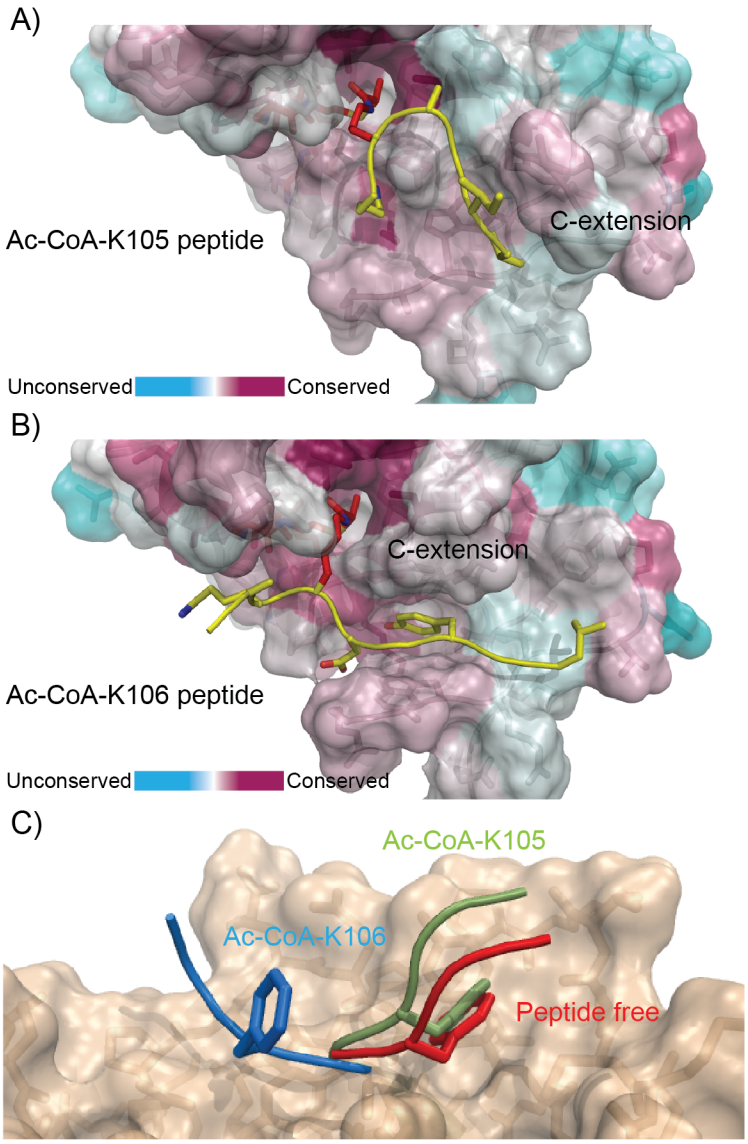


Figure 6.10. C-Extension Moves Upon Binding K106.

A) Orientation of the peptide, Ac-CoA-K105, on the surface of xEco2 ACT. Unconserved and conserved residues are shown in light green and maroon respectively. CoA and the peptide are shown in red and yellow respectively. B) Orientation of the Ac-CoA-K106 peptide on the surface of xEco2 ACT. Colouring is the same as in (A). C) Comparison of the movement of the C-extension in the peptide free xEco2 ACT (red), Ac-CoA-K105 (green) and Ac-CoA-K106 (blue).

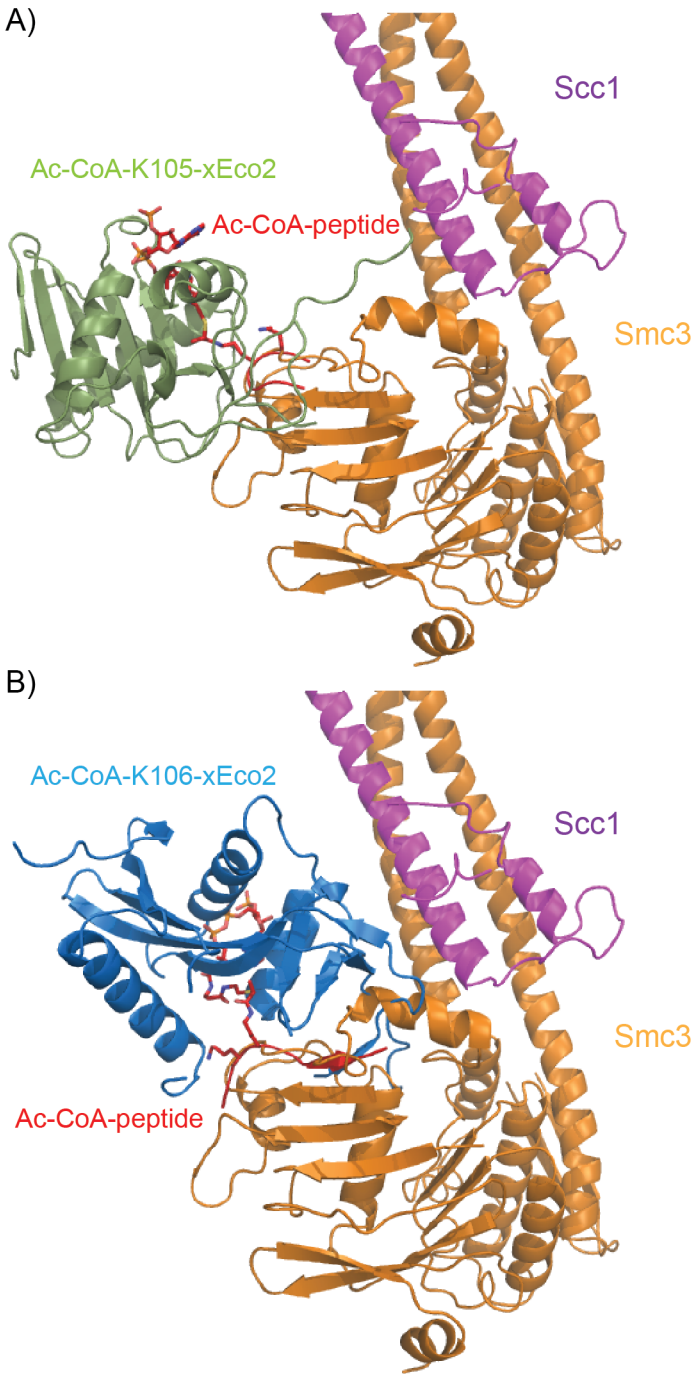


Figure 6.11. Docking of xEco2 Onto Smc3

A) Ac-CoA-K105 (green) structure is docked onto the Smc3(orange)-Scc1N (pink) structure (PDB: 4UX3). In red is shown the peptide-Ac-CoA conjugate. B) Ac-CoA-K106 (blue) structure is docked onto the Smc3 (orange)-Scc1N (pink) structure (PDB:.4UX3). In red is shown the peptide-Ac-CoA conjugate.

6.8 Acetylation Does not Impact Pds5 Binding

During S-phase the tandem lysine motif of Smc3 becomes acetylated. This post-translation modification causes the establishment of sister-chromatid cohesion through antagonising the releasing activity of Wapl and Pds5. Although the role of acetylation is well understood, how it affects the realising activity of Wapl and Pds5 is not. It could be hypothesised that acetylation blocks Wapl and Pds5 binding to Smc3. To test this, the human cohesin head domain (hsCohd) with acetylation mimic (K105Q and K106Q) was recombinantly expressed in insect cells along with human Pds5 (hsPds5). Both proteins were purified using a standard three step procedure and purified proteins were mixed in equi-molar ratios and loaded onto a pre-equilibrated gel filtration column. The results show that mutant hsCohd could still interact with Pds5 shown by the co-migration of Pds5 and hsCohd (Figure 6.12A and B). This results suggested that the acetylation of Smc3 does not change the affinity of Pds5 to cohesin.

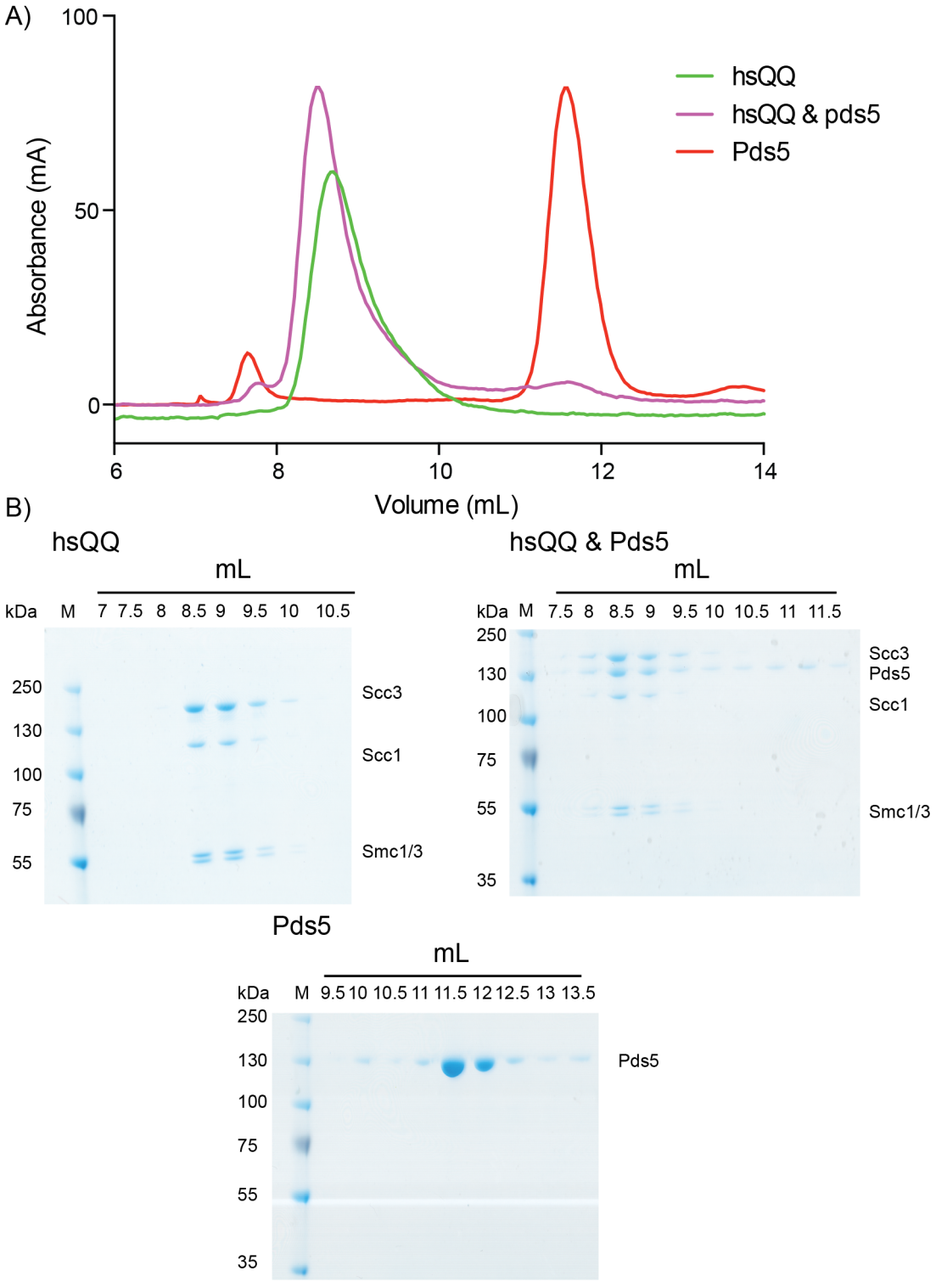


Figure 6.12. Pds5 Binds to Acetylated hsCohd.

A) Gel filtration absorbance profiles for mutant hsCohd (hsQQ) (green), human Pds5 (red) and both hsQQ and pds5 (purple) mixed in an equi-molar ratio on a pre-equilibrated Superdex S200 10/300 GL column (GE Healthcare). B) SDS-PAGE analysis was conducted on the absorbance peaks from the three gel filtration experiments.

6.9 Acetylation Reduces DNA Binding

The release of cohesin from DNA requires the hydrolysis of ATP by the head domain of cohesin, which is stimulation by Wapl, Pds5 and DNA (Murayama and Uhlmann, 2015). The acetylation of Smc3 at K112 and K113 (K105 and K106 in higher eukaryotes) reduces the rate of stimulated ATP hydrolysis (Murayama and Uhlmann, 2015). Therefore, it has been suggested that the tandem lysines of Smc3 could act as DNA sensors (Murayama and Uhlmann, 2015). To better understand the interaction between DNA and cohesin an atomic model of the head domain of cohesin bound to DNA was made using the DNA-Rad50 structure as a model (Rojowska et al., 2014) (Figure 6.13A). The analysis supports the idea that DNA is in close proximity to a range of conserved basic residues found on the coiled-coil of Smc3; more importantly, K112 and K113 of Smc3 would also contact DNA.

To test if the tilted model of DNA binding is correct WT human head domain was purified along with the two Smc3 mutants K105R/K106R (RR) and K105Q/K106Q (QQ) (Figure 6.13B). The purified complexes were titrated against a constant concentration of DNA and an EMSA experiment was carried out (Figure 6.13C). The results indicate that the QQ complex had a reduced affinity for DNA compared to WT and RR head domain complexes. This suggests that acetylation of Smc3 reduces the affinity of the head domain to DNA, which supports the binding model presented in Figure 6.13A.

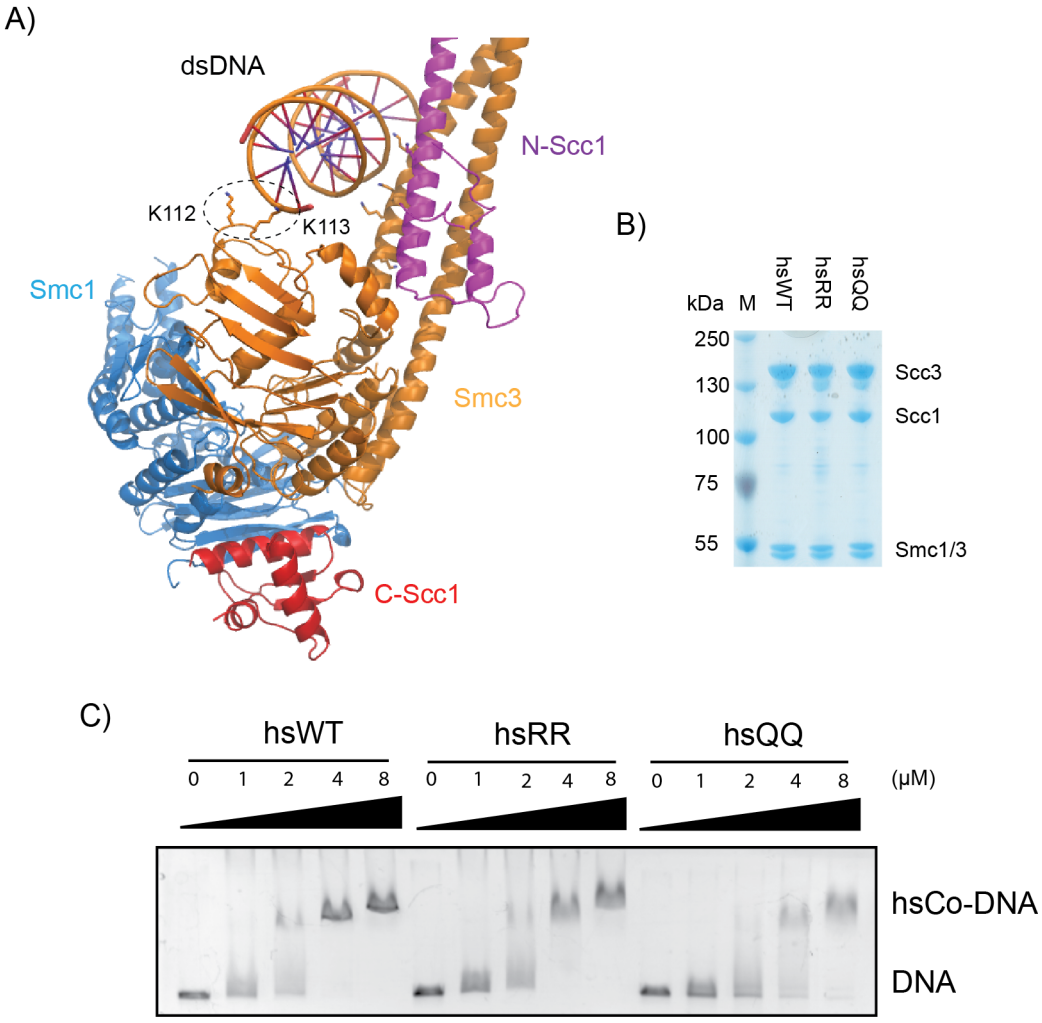


Figure 6.13. Acetylation Reduces DNA Binding of hsCohd.

A) Atomic model for the binding of DNA to the head domain of cohesin. The colouring is as follows: Smc3 (orange), Smc1 (blue), N-Scc1 (pink), C-Scc1 (red) and dsDNA (orange). The tandem lysine motif is indicated in stick representation. B) SDS-PAGE analysis of the three head domain complexes. C) The EMSA experiment conducted with the three human head domain complexes. DNA was kept at a constant concentration and the protein was titrated at the concentration indicated above the gel.

Chapter 7. Discussion

Directly after DNA replication sister-chromatids are held together until the onset of anaphase by a protein complex known as cohesin. The early loss of cohesion between chromatids causes chromosome mis-segregation. Therefore, cells have developed complex regulatory systems to help control cohesion throughout the cell cycle. This 'cohesion cycle' can broadly be split into four stages: loading, anti-establishment, establishment and release.

During G1, cohesin is loaded onto chromatin by a protein complex formed of two proteins called Scc2/4 (Ciosk et al., 2000). At this point, cohesion between sister-chromatids is not established due to the anti-establishment activity of Pds5 and Wapl (Kueng et al., 2006). After S-phase, sister-chromatid cohesion becomes established with the passage of the replication fork due to the acetyltransferase called Eco1 (Ben-Shahar et al., 2008). This protein acetylates Smc3 at K112 and K113, of the cohesin complex, which antagonises the releasing activity of Pds5 and Wapl. Apart from Eco1, numerous other non-essential establishment factors exist; such as Chl1, RFC^{Ctf18}, Tof1, Ctf4 and Mrc1. All of these factors, apart from Chl1 and Ctf4, aid in establishment of sister-chromatid cohesion by Eco1. Chl1 and Ctf4 function in a poorly understood Eco1 independent establishment pathway (Borges et al., 2013). After the establishment of sister-chromatid cohesion, cohesin topologically holds sister-chromatids together until the onset of anaphase; at which point, a protein called separase becomes active (Uhlmann et al., 2000). Separase cleaves the Scc1 subunit of cohesin; breaking the ring structure and allowing the release of chromatin from cohesin's embrace.

In this thesis the aim was to improve our understanding of the establishment factors involved in sister-chromatid cohesion. I focused on two proteins, the key establishment factor Eco1 and the non-essential RFC^{Ctf18} complex. In addition to its role in sister-chromatid cohesion, RFC^{Ctf18} also functions in the DNA damage response checkpoint (DRC). RFC^{Ctf18} is also the only RFC like complex which is formed of seven subunits as it contains the two non-Rfc subunits called Dcc1 and Ctf8. This stable sub-complex takes no part in the loading or unloading of PCNA. Therefore, I hypothesised that Dcc1 and Ctf8 must have a regulatory role. I was able

to solve the structure of the Dcc1-Ctf8 sub-complex with the C-terminus of Ctf18, which gave novel insights into the role of Dcc1. Furthermore, I was able to solve the structure of the C-terminal acetyltransferase domain (ACT) of xEco2 bound to its two conical substrate peptides. This revealed structural insights into how Eco1 is able to target and acetylate both tandem lysine residues of Smc3. In addition, biochemical experiments were conducted to determine the effect acetylation had on the head domain of cohesin. In this chapter the discussion will firstly focus on the results relating to RFC^{Ctf18}.

7.1 Structure of the $Ctf18^{c}$-Dcc1-Ctf8 Heterotrimer

The RFC^{Ctf18} complex is formed of seven subunits and has roles in both sister-chromatid cohesion and the DRC (Mayer et al., 2001; Pan et al., 2006). The core RFC like complex is formed from Ctf18 and Rfc2-5 (Mayer et al., 2001). Unique to this complex is the additional requirement for two non-Rfc subunits called Dcc1 and Ctf8. These proteins form a stable sub-complex *in vitro* and cause the same genetic effect as a Ctf18 deletion. Previous research has revealed that Dcc1 and Ctf8 take no part in the loading or unloading of PCNA from chromatin, which suggests this sub-complex has a regulatory role (Bylund and Burgers, 2005). To further understand the function of the Dcc1-Ctf8 sub-complex I crystallised the C-terminal region of Dcc1 alone; and Dcc1 and Ctf8 bound to the C-terminal region of Ctf18 ($Ctf18^{C}$).

The results, unexpectedly, revealed that the C-terminal region of Dcc1 is formed of three tandem winged helix (WH) domains, which I have called WH1, WH2 and WH3 in this thesis. The organisation of these three WH domains in a tandem conformation is novel and to my knowledge not observed in any previously solved structures. Usually, WH domains are involved in protein-protein interaction or DNA-protein interactions (Harami et al., 2013), which suggests a possible function for these domains. To test whether Dcc1 could interact with DNA an EMSA experiment was conducted, which indicated that Dcc1 was able to bind to both ssDNA and dsDNA. Mutating conserved basic residues in the C-terminal region of Dcc1 revealed that the binding site for ssDNA and dsDNA overlapped as mutations blocked both ssDNA and dsDNA binding to the same extent. The majority of the affinity for Dcc1 to both

ssDNA and dsDNA was given by WH3. However, further analysis showed that both WH1 and WH2 aided the binding of ssDNA but not dsDNA.

The structure of the Ctf18^{c}-Dcc1-Ctf8 heterotrimer revealed that Dcc1 and Ctf8 interact to produce a large dimer interface, which forms a fold resembling the triple β-barrel structures previously identified in the three transcription factors Rap30/74 (Gaiser et al., 2000), Sfc1/7 (Taylor et al., 2013) and A39/A34.5 (Geiger et al., 2010); and the heterotrimer RNase H2 formed of RNaseH2A, RNase H2B and RNase H2C (Figiel et al., 2011). Much like Ctf18^{c}, the C-terminal region of RNase H2A binds across the dimer formed from RNase H2B and C. The structural similarities of Ctf18^{c}-Dcc1-Ctf8 with the transcription factors also extends to them sharing a similar modular organisation. These transcription factors also use this dimerisation domain to link a larger complex to DNA binding motifs; most of which are able to bind to both ssDNA and dsDNA. This raises the interesting possibility that Dcc1 and Ctf8 evolved from a distant relative of one of these transcription factors. Previous work supports the model for Ctf18 binding to the Dcc1-Ctf8 heterodimer like an extended peptide as mutating both W736 and W740 causes a disruption in the formation of the heterotrimer (García-Rodríguez et al., 2015). These key tryptophan residues bind as a partially buried α-helix in the first β-barrel of the heterodimer interface.

RFC like complexes usually interact with a range of DNA substrates through a variety of Rfc subunits (Gomes et al., 2001). The addition of an extra DNA binding domain at the C-terminal region of Dcc1 in the RFC^{Ctf18} complex raises the intriguing possibility that Dcc1 plays a vital role in the regulation of this complex. This hypothesis was confirmed with the discovery that deleting WH3 of Dcc1 drastically reduced the amount of Ctf18 at stalled replication forks. However, the results did not fully replicate a *dcc1Δ* as in this strain less Ctf18 was present at stalled replication forks compared to the WH3 deletion. This results suggests that RFC^{Ctf18} could be recruited by multiple mechanisms. It is interesting that WH3 also supports the interaction with Pol2, a subunit of polymerase-ε. This makes it difficult to conclude whether the recruitment of RFC^{Ctf18} by Dcc1 is through direct interaction with DNA or through binding Pol2. Furthermore, the *in vivo* data indicated that deleting WH3 caused a delay in Rad53 phosphorylation and a reduced cell survival in HU

supporting the hypothesis that the WH domains of Dcc1 play a regulatory role on the RFC^{Ctf18} complex.

The availability of a structure of PCNA bound to RFC^{Rfc1} (Bowman et al., 2004) permitted the docking of $Ctf18^{C}$-Dcc1-Ctf8 to Rfc1, which created a speculative model for the full RFC^{Ctf18} complex (Figure 7.1). The model was made based on the bacteria γ clamp loader complex, which is also formed of seven subunits (Simonetta et al., 2009). The core complex is formed from three γ subunits, one δ subunit and one δ' subunit. It also has a stable sub-complex associated with it, which is formed from the two subunits, ψ and χ. This stable sub-complex does not show any structural similarities to the Dcc1-Ctf8 heterodimer. However, like RFC^{Ctf18} the core γ complex is able to load the bacterial clamp onto DNA without the accessory subunits. The ψ-χ sub-complex is involved in recruiting the γ complex to DNA through its interaction with single-stranded binding protein (SSB), which is a homologue to eukaryotic RPA (Gulbis et al., 2004). It is therefore likely that the Dcc1-Ctf8 sub-complex plays a similar role in recruiting RFC^{Ctf18} to chromatin through its interaction with Pol2 and DNA.

In the γ complex the accessory subunits are believed to be positioned behind the collar region of the core complex (Simonetta et al., 2009). Therefore, it is likely that the Dcc1-Ctf8 sub-complex will be in the same orientation. This places the WH domains of Dcc1 close to the ssDNA leaving the collar region of the core RFC complex (Figure 7.1A and B). If the Dcc1-Ctf8 sub-complex was flipped 180 ° it would interfere with the loading or unloading of PCNA making this geometry unlikely.

Pol2 binds to the surface of the $Ctf18^{C}$-Dcc1-Ctf8 complex (Figure 7.1); however, it is difficult to predict how the full eight subunit complex will look as the analysis of the interaction by glycerol gradients indicates that Pol2 binds to the β-barrel region and to the WH domains. Placing the Dcc1-Ctf8 sub-complex behind the collar region will likely position the polymerase in close proximity to the primer-template junction.

The position of RFC^{Ctf18} at stalled replication forks along with the suggestion that the DRC is activated by a build-up of ssDNA indicates that the target for the WH domains

of Dcc1 is likely to be primed-ssDNA or ssDNA. However, the results presented in this thesis imply that primed-ssDNA or ssDNA did not have a higher affinity than dsDNA. It is possible that substrate specificity is given through its interaction with polymerase-ε.

In conclusion, I have solved the structure of the Ctf18^{C}-Dcc1-Ctf8 heterotrimer, which has revealed that the C-terminal region of Dcc1 contains three WH domains. These WH domains function in binding both ssDNA and dsDNA and are involved in recruiting the complex to stalled replication forks.

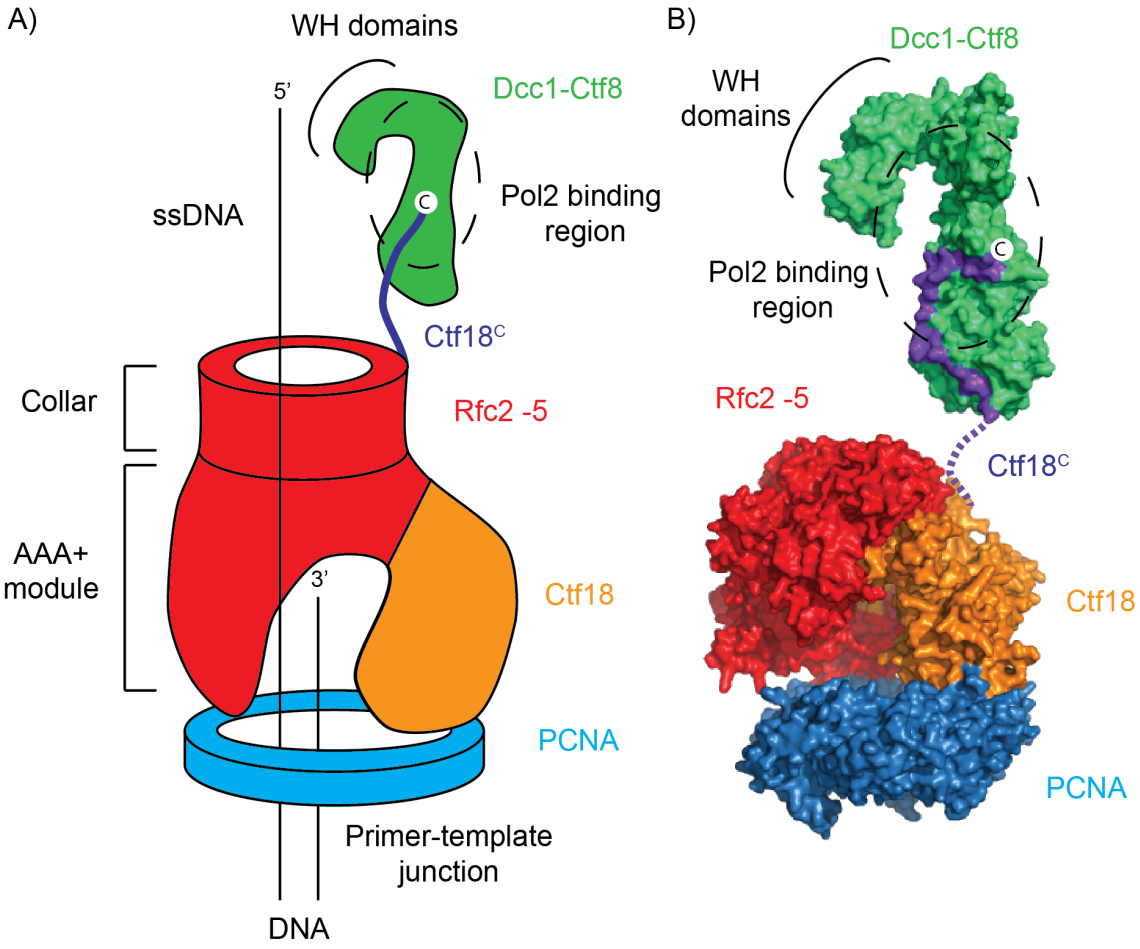


Figure 7.1. Predicted Structure of RFC^{Ctf18}.

A) Model for the RFC^{Ctf18} complex. The colouring is as follows; PCNA (blue), Ctf18 (orange), Rfc2-5 (Red), Ctf18^{C} (purple), Dcc1-Ctf8 (green) and DNA (black). The position of the WH domains of Dcc1 are labelled in the figure. The black dotted line indicates the Pol2 binding region. B) Model for the RFC^{Ctf18} complex based on the RFC^{Rfc1} crystal structure and the Ctf18^{C}-Dcc1-Ctf8 crystal structure. Colouring is the same as in (A).

7.2 Structural and Biochemical Analysis of Eco1

Eco1 is the key establishment factor in sister-chromatid cohesion. It was first discovered to be an essential gene which did not form part of the core cohesin complex (Toth et al., 1999). Later it was found to contain a C-terminal GCN5-related N-acetyltransferases (GNAT) domain and an N-terminal C_2H_2 zinc finger motif (ZnF) (Ivanov et al., 2002). The N-terminus of Eco1 is divergent in both length and composition and is thought to provide species specific roles. During S-phase, Eco1 is recruited to the replication fork through its interaction with PCNA via a conserved PCNA interaction motif or PIP box (Moldovan et al., 2006). Eco1 is then able to acetylate cohesin on K112 and K113 (K105 and K106 in higher eukaryotes) of the Smc3 subunit, which antagonises the action of Pds5 and Wapl (Ben-Shahar et al., 2008). In addition, Eco1 is able to acetylate Scc1 at K84 during a DSB and auto-acetylate itself *in vitro* (Heidinger-Pauli et al., 2008; Ivanov and Nasmyth, 2005).

To gain further insight into how Eco1 is able to target the tandem lysine motif I firstly designed a series of *in vitro* assays coupled to mass spectrometry (MS). The results from these experiments indicated that K112 acetylation is faster and, possibly, facilitates the slower acetylation of K113. In addition, I identified a bi-directional peptide motif, which showed the target lysine was flanked by an aliphatic and basic/polar residue. To further support the above MS analysis, I solved the structure xEco2 ACT domain bound with two conjugated substrate peptides. Finally, the role of acetylation was investigated with the *in vitro* reconstitution of the head domain of cohesin.

The discovery that Eco1 acetylates K112 before K113 could suggests that the basic side chain of K112 needs to be neutralised by an acetyl group before Eco1 can bind and acetylate K113. However, previous genetics would disagree with this argument because of two observations (Zhang et al., 2008). Firstly, K113 is more important *in vivo* as the mutant strain K112R is completely viable but K113R has a reduced survival rate and an increase in chromosome mis-segregation. Therefore, if K112 acetylation was a pre-requisite of K113 acetylation you would expect there to be a large phenotype in the K112R strain, as in these cells, K113 could not become acetylated. This is not the case and therefore it is likely that the kinetics of the two

acetylation steps are different. Potentially, the acetylation of K112 occurs at as a faster rate than K113 acetylation, which is supported by the two peptide structures. For instance, upon binding K113 (Ac-CoA-K106) the C-extension of Eco1 is displaced by almost 180 °, which could be kinetically unfavourable; however, when Eco1 is bound to K112 (Ac-CoA-K105) the C-extension adopts the same conformation as in the peptide free structure making this substrate more favourable. This is also supported by the observation that the peptide density for Ac-CoA-K106 is much stronger, which could indicate that the peptide has a slow on/off rate.

The results from the peptide motif MS experiments indicate that Eco1 was able to acetylate a consensus motif which had a lysine flanked by an aliphatic residue (B/Z) and an acidic/polar (ϕ) residue (B/Z-AcK-ϕ). This result was supported by the structure of Ac-CoA-K106, which showed that D107 at position +1 (P+1) interacted with the indole side chain of W623 and formed a salt bridge with R621. In addition, at P-1 from the unconjugated lysine (K105) interacted through its aliphatic moieties with L562 and F564 of xEco2. Previous evidence has suggested that for Scc1 to become acetylated at K84, Chk-1 is required to phosphorylate S83 (Heidinger-Pauli et al., 2008), which could provide an acidic mimic to support strong Eco1 binding.

A proportion of the peptides identified could not be classed into the above bi-directional motif and were grouped as having ‘no direction’. These peptides were generally flanked by a proline or arginine which supports the β-hairpin conformation of the peptide observed in the structure of Ac-CoA-K105 and Smc3-Scc1N. Furthermore, in the peptide Ac-CoA-K105 structure K106 at P+1 salt bridges with D677 showing the importance of having a lysine residue which is not acetylated at this position. This supports the analysis from MS data and could be the reasons K112 acetylation occurs first. In addition, it provides support for having a basic residue at P+1 in the peptide motif with no direction.

Eco1 was able to acetylate a range of cohesin targets, *in vitro*, albeit at a high concentration. It raises the intriguing possibility that apart from acetylating Smc3 during S-phase and Scc1 in response to DSB, Eco1 may be able to acetylate other cohesin related targets. The peptide motif and structures presented in this thesis could provide a rationale for investigating this further.

The ACT domain of xEco2 looks similar to other GNAT family proteins. However, between the N- and C-lobe there is an Eco1 specific insertion which forms a β-hairpin structure. The two structures solved in this thesis revealed that xEco2 formed two different conformations when bound to different conjugated peptides. Multiple contacts were made by both peptides with large movements being observed in the C-extension in the Ac-CoA-K106 structure (Figure 7.2).

The effect, *in vivo*, of acetylation on Smc3 at K112 and K113 causes the establishment of sister-chromatid cohesion through antagonising the action of Pds5 and Wapl. However, the biochemical function of acetylation is still relatively unknown. It has been suggested in previous studies that K112 and K113 act as DNA sensors to trigger the loading or release of DNA through the entry and exit gate respectively (Murayama and Uhlmann, 2015). To test this directly, the head domain of cohesin was recombinantly purified with acetylation mimics. Using these complexes, the DNA binding to dsDNA was tested. The results indicated that acetylation reduced the affinity of the head domain to DNA. This suggested that acetylation of the tandem lysine motif would block DNA stimulated hydrolysis, which agrees with the conclusions drawn from previous studies.

The greater importance of K113 acetylation over K112 acetylation most likely relates to K113 being more significant in contacting DNA. Acetylation of K112 and K113 did not completely abolish DNA binding, which is unsurprising as there are numerous other surface exposed basic residues present around the proposed DNA binding site. More recently other binding models of Rad50 have been suggested rather than the tilted model presented in Figure 6.13A (Rojowska et al., 2014). In these conformation DNA would not contact the tandem lysine motif. However, I propose that the tilted DNA binding is an intermediate step and the head domain can probably interact with DNA in multiple protein-DNA geometries.

In conclusion, I have discovered that K112 acetylation occurs before the acetylation of K113. I have also solved the structure of xEco2 bound to two conjugated peptides which has shown the importance of both the conserved C-extension and the β-

hairpin. Finally, I have shown that acetylation blocks DNA binding to the head domain in a tilted conformation, which inhibits the DNA induced hydrolysis of ATP.

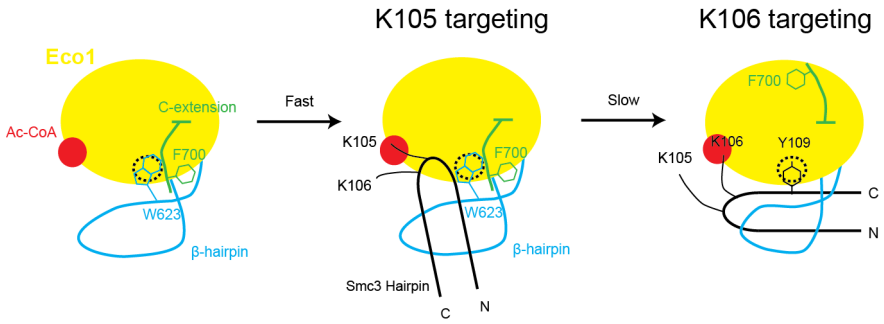


Figure 7.2. Conformational Changes in Eco1 Upon Substrate Binding.

7.3 Future Work

7.3.1 Further work on the RFC^{Ctf18} complex

In this thesis I have presented work which improves our understanding of the role the Ctf8-Dcc1 sub-complex plays in regulating RFC^{Ctf18}. However, to gain further insights, two structural approaches could be taken. In the first instance, it would be useful to obtain a complete structural model for RFC^{Ctf18}. This could be done with the use of insect cells to express the seven-subunit complex. Once the full complex is expressed, structural studies could be carried out using electron microscopy (EM) or by X-ray crystallography. Hopefully, this will provide further support to the pseudo-atomic structure presented in Figure 7.1. Additionally, to gain further information about the binding mechanisms of Dcc1 one could attempt to obtain a crystal structure of the C-terminal region of Dcc1 bound to DNA. This would potentially give additional insights into the type of substrate Dcc1 interacts with.

In the work presented in Chapter 5 it was clearly shown that WH3 of Dcc1 binds to DNA and is involved in recruiting the complex to chromatin. However, more work is needed to clarify the function of WH2 and WH1. It could be possible that they serve as additional DNA binding sites to support WH3 although it is also conceivable that WH1 and WH2 carry out other roles.

7.3.2 Further work on the role of Eco1

To gain further insights into the role of Eco1 future studies could further investigate the role of K112 acetylation. It would be interesting to observe if *in vitro* it would be possible for K113 acetylation to occur with the mutation K112R, albeit at a reduced rate. This would link the results presented in this thesis with the previous observed genetic observations (Zhang et al., 2008).

Further work could also be conducted into the role of the N-terminus of Eco1 as a conserved ZnF domain is present in all Eco1 homologs; it would be interesting to assign a biological function to this motif. Structural studies could be carried out on the N-terminus of Eco1 which may aid in our understanding of the function of this region.

The model for the role of acetylation presented in Chapter 6 implies that K112 and K113 regulate ATP hydrolysis of the head domain by sensing DNA. To further clarify this the structure of Smc3-Scc1N could be solved with DNA to support the tilted binding model of the Rad50-DNA complex (Rojowska et al., 2014).

7.4 Summary of Findings Presented in This Thesis

In summary, I solved the structure of the C-terminus of Dcc1 along with the Ctf18-Dcc1-Ctf8 heterotrimer, which forms part of the RFC^{Ctf18} complex. I have shown that the C-terminus of Dcc1 contains three tandem WH domains which can bind to both ssDNA and dsDNA. Furthermore, I have revealed that WH3 of Dcc1 is important for recruiting the complex to stalled replication forks. I have also solved the structure of xEco2 ACT domain in two peptide bound structures, which has indicated that Eco1 is able to bind the tandem lysine motif through different conformations. In addition, I have presented results which show that the acetylation of Smc3 reduces the affinity of DNA to the head domain of cohesin. This has provided evidence that the tandem lysines motif is important in sensing DNA. Finally, I have revealed that the kinetics of the acetylation K112 and K113 is different as K112 acetylation occurs first.

7.5 List of Publications From This Thesis

Samora, C. P et al., (2016). Ctf4 Links DNA Replication with Sister Chromatid Cohesion Establishment by Recruiting the Chl1 Helicase to the Replisome. *Molecular Cell*. http://doi.org/10.1016/j.molcel.2016.05.036

Chao, W. C. H., Murayama, Y., Muñoz, S., Jones, A. W., Wade, B. O., Purkiss, A. G., et al. (2017). Structure of the Cohesin Loader Scc2. *Nature Communications*, *8*, 13952. http://doi.org/10.1038/ncomms13952

Wade, B. O., Liu, H. W., Samora, C. P., Uhlmann, F., & Singleton, M. R. (2017). Structural Studies of RFC^{Ctf18} Reveal a Novel Chromatin Recruitment Role for Dcc1. *EMBO Reports*, e201642825. http://doi.org/10.15252/embr.201642825

Chao W. C. H & Wade B. O, et al., (awaiting publication) Structural Basis of Eco1-Mediated Cohesin Acetylation.

Appendix

Appendix 1. List Restriction Free (RF) Primers Used in this Thesis.

For the forward (Primer 1) and the reverse primers (Primer 2) the sequence corresponding to the vector sequence is in orange and the protein codding sequence is in blue. The primer sequence runs in the 5' to 3' direction.

Construct	Vector	Primer 1	Primer 2
Dcc1	pETDuet	AATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGTCCATCAACCTACATTCCG	CGACTTAAGCATTATGCGGCCGCCTACCTGCTCGTGACCAC
Ctf8	pCDFDuet	AATAATTTTGTTTAACTTTAATAAGGAGATATACCATGCCTAGTGTAGATATAGATGCTA	ACAATACGATTACTTTCTGTTCGACTTAAGCATTATTACATAATAGGTAGAGGCCTATCC
Dcc1 (90 - 380)	pET28a	GCGAAAACCTGTATTTTCAGAGCGGAGGACTGTCCAAGCCGTAC	TCAGCTTCCTTTCGGGCTTTGTTACTACCTGCTCGTGACCAC
Ctf18 (666 - 741)	pET28a	CGAAAACCTGTATTTTCAGAGCGGATCCAAAGTCAAAACTGGGTTAAATTCTT	TCAGCTTCCTTTCGGGCTTTGTTATTATTCCCACAGGTTATTCCAAGT
xEco2 (523 - 702)	pET28a	CACAGCGAAAACCTGTATTTTCAGAGCATGAAGAAAGAGCGAGTTATTACAGAGT	TCAGCTTCCTTTCGGGCTTTGTTACTAACTGACAAAATTATACACCAGG
Eco1	pET28a	CGAAAACCTGTATTTTCAGAGCGGATCCATGAAAGCTAGGAAATCGCAGA	TCAGCTTCCTTTCGGGCTTTGTTATCATATGTATACCGGCAATAGTAAC
Dcc1	pET28a	GCGAAAACCTGTATTTTCAGAGCGGAATGTCCATCAACCTACATTCCG	TCAGCTTCCTTTCGGGCTTTGTTACTACCTGCTCGTGACCAC
Pol2 (1 - 533)	pET28a	CGAAAACCTGTATTTTCAGAGCGGATCCATGATGTTTGGCAAGAAAAAAAACA	TCAGCTTCCTTTCGGGCTTTGTTAATTTGGTAGAAGAATATTATGTTGA
Dcc1ΔWH2-3	pETDuet	AATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGTCCATCAACCTACATTCCG	ACAATACGATTACTTTCTGTTCGACTTAAGCATTACAAGCGCCACGTATTGTTT
Dcc1ΔWH3	pETDuet	CGAAAACCTGTATTTTCAGAGCGGATCCATGTCCATCAACCTACATTCCG	TCAGCTTCCTTTCGGGCTTTGTTATTTCGCTATATACTGGACAGTCTTA

Appendix 2. List Primers Used for Mutagenic Cloning in This Thesis.

Protein	Mutation	Primer 1	Primer 2
Dcc1	K326A	CCAATGGACCCCGCAGA ACGGTTTAAAGTCC	GGACTTTAAACCGTTCT GCGGGGTCCATTGG
Dcc1	K357A	CAAGAGGTATGGCAAT AGACAGTTTCATCATG	CATGATGAAACTGTCTA TTGCCATACCTCTTG
Dcc1	K364A	GACAGTTTCATCATGGC GTATGCCCGCCG	CGGCGGGCATACGCCA TGATGAAACTGTC
Dcc1	R367A	CATGAAGTATGCCGCCC GTAAAAGACTGGGC	GCCCAGTCTTTTACGGG CGGCATACTTCATG
Dcc1	R380A	AAAGACCGTGGTCACG AGCGCGTAGAGCTTGC	GCAAGCTCTACGCGCTC GTGACCACGGTCTTT

Appendix 3. List of Acetylated Peptides From *in vitro* Acetylation Assay.

Forward	
Gene names	**Sequence window**
PDS5	TSDQLISTNELLDRLKALHEELASLDQDNTD
PDS5	TESIPQIIATREDISKELNQALAKTFIDSDP
PDS5	ATREDISKELNQALAKTFIDSDPRVRRTSVM
PDS5	TSVMIFNKVPVTEIWKNITNKAIYTSLLHLA
PDS5	GLSDSTKAIDALETIKQFNDERIFYLLNACV
PDS5	RRILDDISKVNPTLFKDQIRTLKTIIKDLDD
PDS5	PTLFKDQIRTLKTIIKDLDDPDAEKNDNLSL
PDS5	TLKKIKIRILPLDLQKDKYFTSHIIVLMEIF
PDS5	LIKEVLLSNQVVGDSKKEIDWVEDSLLSDTK
PDS5	QVLKLARISNLNNFIKPSDIIKLINLVEDES
PDS5	DKLVEDEIDEEEGPQKEEAPKKHRPYGQKMY
RAD61	NVKIITSSDTSMAFMKDEKLSAFNFLDGSKA
RAD61	SKDKKYGKFRTILINKNKENEIMGEEVDQKA
RAD61	AEKEGLTSTNHYNELKNMGDTIKYQDDIEFL
RAD61	SKSNDNTTVPINEYFKKLLNLSLMIINDEEF
RAD61	SDFPPSFDNFSIELSKDDGKIRTKKNKHIKK
RAD61	LSLDIIKRISILASNKDLFSRHVKTFIPLLE
RAD61	ISILASNKDLFSRHVKTFIPLLEKLITASEF
SCC1	GNIDTITDAMTESQPKQTGTRRNSKLLNTKS
SCC1	LPDPILKNFLSYESLKKRKIHNGREGSIEEP
SCC2	CNQMKDIAPENIDLLKNEYKKQEEFLFNIVE
SCC2	TKDAKEQIITVDNELKKILEQIKDGGLGPEL

SCC2	SLLAKDKIKLRSTAIKCLSMLASKDKVILSN
SCC2	RNTGVNSTLSSNSILKKKLLKTNRVEFANDG
SCC2	RFLDNILQLCLLRDLKNSLVAIRLLKLILKF
SCC3	SIIDNIVKLCFVHRYKDVSDLIRSESMLHLS
SCC3	REIIEKVRDYENSISKDLRSVWKPIAAIIGR
SCC4	RKFLPKVYSTTQKLIKNIAAGGVSMNELDSR
SCC4	GKDVALTNAKLEALVKQITSVKQ_____
SMC1	___MGRLVGLELSNFKSYRGVTKVGFGESNF
SMC1	FVLGVRSNHLRSNILKDLIYRGVLNDENSDD
SMC1	SIKNRRRIHGELKTYKEGINKNEEYRKQLDK
SMC1	LTDKLSALNSEISSLKGKINNEMKSLQRSKS
SMC1	TDRLSTTQRRYEKAQKDLENAQVEMKSLEEQ
SMC3	SRGDDEVTIRRTVGLKKDDYQLNDRNVTKGD
SMC3	RGDDEVTIRRTVGLKKDDYQLNDRNVTKGDI
Reverse	
ECO1	RIWVCRTARKLGIATKLIDVARENIVYGEVI
ECO1	PRYQVAWSQPTDSGGKLASKYNGIMHKSGKL
PDS5	_____MAKGAVTKLKFNSPIISTSDQLI
PDS5	LASLDQDNTDLTGLDKYRDALVSRKLLKHKD
PDS5	DKRVHRLLTVLSHFDKKAFTSFFAFNARQIK
PDS5	IPFLTFKNCYNELVSKLQTPGLFKKYNISTG
PDS5	AKQLDLKRRILDDISKVNPTLFKDQIRTLKT
PDS5	TKLYDFAVESKPEITKYATKLIALSPKAEET
PDS5	DFAVESKPEITKYATKLIALSPKAEETLKKI
PDS5	KKEIDWVEDSLLSDTKYSAIGNKVFTLKLFT
PDS5	EDSLLSDTKYSAIGNKVFTLKLFTNKLRSIA
PDS5	SAIGNKVFTLKLFTNKLRSIAPDVPRDELAE
PDS5	APDVPRDELAESFTEKTMKLFFYLIASGGEL
PDS5	EFNKEFYPTPSNYQTKLRCVAGIQVLKLARI
PDS5	SLLYYLSERVKNYQDKLVEDEIDEEEGPQKE
PDS5	AQLSFKTYIPESLTEKIQNNIKAKIGRILHT
RAD61	IITSSDTSMAFMKDEKLSAFNFLDGSKASKR
RAD61	DGSKASKRKRRRTYQKHDANITSSIEPDVQD
RAD61	EFILKLPRADDDILNKMLENEMKMDDSIENN
RAD61	NKNKENEIMGEEVDQKANTLSLNNADNSNAE
RAD61	PSFDNFSIELSKDDGKIRTKKNKHIKKLSHL
RAD61	KDDGKIRTKKNKHIKKLSHLNFEDFLRKTQF
RAD61	LNICSRENSRLKLDGKLWYDMKTIFVKMIRD
SCC1	KDTLTKISMLFKTSQKMTSTVNRLNTVTRVH
SCC1	TIPVGLMAQENSMERKVQGAAPWDTSLEVGR

SCC1	MTESQPKQTGTRRNSKLLNTKSIQIDEETEN
SCC1	IDAKTRNEQTTIQTEKVRPTPGEVASKAIVQ
SCC1	PTPGEVASKAIVQMAKILRKELSEEKEVIFT
SCC2	IEKKSEIVSRPEAKHKLESVTSNAGNLSFND
SCC2	DDSIMVRKHVLRINEKMYDETNDIVTKVYVI
SCC2	LVQKIVELNSDDTNEKNSIVDKQNFLNLLAK
SCC2	INKANNEEAAIVVDGKLQRLIYLSTGFARFC
SCC2	LYEHITKCLLVLSKDKITHVIRRVAVKNLTK
SCC4	CEFLLLHDLPLMRDSKFHYKIALRNCNELVQ
SCC4	NCYDEKGNFSRKFLPKVYSTTQKLIKNIAAG
SCC4	NFSRKFLPKVYSTTQKLIKNIAAGGVSMNEL
SCC4	EQTLLKGAVVTTESPKLGPSPGYVRLLQAMK
SCC4	CYTVQAARVSRCSGDKQGELVEQCNKVWLQV
SCC4	CTVAMRGKDVALTNAKLEALVKQITSVKQ__
SMC1	SSFVKESAVISKQKSKLDYIFKDKEKLVSDL
SMC1	KLNENDLKTYNCLHEKYLTEGGSILEEKIAV
SMC1	NEKNALHTERLHELKKLQSDIESANNQEYDL
No direction	
ECO1	AVGIIIIENLYGGNGKTSSRGRWMVYDSRRL
ECO1	VYDSRRLVQNVYPDFKIGISRIWVCRTARKL
ECO1	DSGGKLASKYNGIMHKSGKLLLPVYI_____
PDS5	KRVHRLLTVLSHFDKKAFTSFFAFNARQIKI
PDS5	DTFFFTKLYDFAVESKPEITKYATKLIALSP
PDS5	PEITKYATKLIALSPKAEETLKKIKIRILPL
PDS5	NYQTKLRCVAGIQVLKLARISNLNNFIKPSD
PDS5	DEIDEEEGPQKEEAPKKHRPYGQKMYIIGEL
PDS5	TYIPESLTEKIQNNIKAKIGRILHTSQTQRQ
PDS5	RLQKRLLAHENNESQKKKKKVHHARSQADDE
PDS5	SDDDSYSPSNKNETKKGHENIVMKKLRVRKE
RAD61	SFDGANEKPSSQLDSKRNDQNVKIITSSDTS
RAD61	KPSSQLDSKRNDQNVKIITSSDTSMAFMKDE
RAD61	EKLSAFNFLDGSKASKRKRRRTYQKHDANIT
RAD61	ANTLSLNNADNSNAEKEGLTSTNHYNELKNM
RAD61	STNHYNELKNMGDTIKYQDDIEFLLSNSKSN
RAD61	TIKYQDDIEFLLSNSKSNDNTTVPINEYFKK
RAD61	EFFQYAKRYFKKEIIKLSFAQFRSDFPELIL
SCC1	RSNNLLTPQPTNFTTKRLWSEITESMSYLPD
SCC1	DVLKSQANTEPENITKREASRGFFDILSLAT
SCC1	GLSQTEAFGNIKIDAKPALFERFINA_____
SCC2	TSNAGNLSFNDNSSNKKTKTSTGVTMTQANL

SCC2	AGNLSFNDNSSNKKTKTSTGVTMTQANLAEQ
SCC2	RLIYLSTGFARFCFPKPSNDKIAFLQEGETL
SCC2	IVEESELKNKQLPAKKPDISKFSAQLENIEQ
SCC3	___MTAVRRSTRIRTKSQVIEEDYDDEQNTS
SCC4	NFFDTNKQSLVTNEGKGCVIKIMPRIALKVE

Reference List

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H., 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. 66, 213–221. doi:10.1107/S0907444909052925

Anderson, D.E., 2002. Condensin and cohesin display different arm conformations with characteristic hinge angles. The Journal of Cell Biology 156, 419–424. doi:10.1083/jcb.200111002

Aparicio, O.M., Stout, A.M., Bell, S.P., 1999. Differential assembly of Cdc45p and DNA polymerases at early and late origins of DNA replication. Proc. Natl. Acad. Sci. U.S.A. 96, 9130–9135.

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., Ben-Tal, N., 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Research 44, W344–W350. doi:10.1093/nar/gkw408

Barber, T.D., McManus, K., Yuen, K.W.Y., Reis, M., Parmigiani, G., Shen, D., Barrett, I., Nouhi, Y., Spencer, F., Markowitz, S., Velculescu, V.E., Kinzler, K.W., Vogelstein, B., Lengauer, C., Hieter, P., 2008. Chromatid cohesion defects may underlie chromosome instability in human colorectal cancers. Proc. Natl. Acad. Sci. U.S.A. 105, 3443–3448. doi:10.1073/pnas.0712384105

Beckouët, F., Srinivasan, M., Roig, M.B., Chan, K.-L., Scheinost, J.C., Batty, P., Bin Hu, Petela, N., Gligoris, T., Smith, A.C., Strmecki, L., Rowland, B.D., Nasmyth, K., 2016. Releasing Activity Disengages Cohesin's Smc3/Scc1 Interface in a Process Blocked by Acetylation. MOLCEL 61, 563–574. doi:10.1016/j.molcel.2016.01.026

Ben-Shahar, T.R., Heeger, S., Lehane, C., East, P., Flynn, H., Skehel, M., Uhlmann, F., 2008. Eco1-Dependent Cohesin Acetylation During Establishment of Sister Chromatid Cohesion. Science 321, 563–566. doi:10.1126/science.1157774

Bernard, P., Drogat, J., Maure, J.-F., Dheur, S., Vaur, S., Genier, S., Javerzat, J.-P., 2006. A Screen for Cohesion Mutants Uncovers Ssl3, the Fission Yeast Counterpart of the Cohesin Loading Factor Scc4. Current Biology 16, 875–881. doi:10.1016/j.cub.2006.03.037

Boersema, P., Raijmakers R., Lemeer L., Mohammed S., Heck A.J.R., 2009. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nature Protocols *4*(4), 484–494. http://doi.org/10.1038/nprot.2009.21

Borges, V., Lehane, C., Lopez-Serra, L., Flynn, H., Skehel, M., Ben-Shahar, T.R., Uhlmann, F., 2010. Hos1 Deacetylates Smc3 to Close the Cohesin Acetylation Cycle. Mol. Cell 39, 677–688. doi:10.1016/j.molcel.2010.08.009

Borges, V., Smith, D.J., Whitehouse, I., Uhlmann, F., 2013. An Eco1-independent sister chromatid cohesion establishment pathway in S. cerevisiae. Chromosoma. doi:10.1007/s00412-013-0396-y

Bowman, G.D., O'Donnell, M., Kuriyan, J., 2004. Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. Nature 429, 724–730. doi:10.1038/nature02585

BRAGG, W.H., 1913. X-rays and Crystals. Nature 90, 572–572. doi:10.1038/090572c0

Brands, A., Skibbens, R.V., 2005. Ctf7p/Eco1p exhibits acetyltransferase activity--but does it matter? Curr. Biol. 15, R50–1. doi:10.1016/j.cub.2004.12.052

Branzei, D., Foiani, M., 2009. The checkpoint response to replication stress. DNA Repair 8, 1038–1046. doi:10.1016/j.dnarep.2009.04.014

Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., Paciorek, W., 2003. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. Acta Crystallogr. D Biol. Crystallogr. 59, 2023–2030.

Brünger, A.T., 1992. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature 355, 472–475. doi:10.1038/355472a0

Bylund, G.O., Burgers, P.M.J., 2005. Replication Protein A-Directed Unloading of PCNA by the Ctf18 Cohesion Establishment Complex. Mol. Cell. Biol. 25, 5445–5455. doi:10.1128/MCB.25.13.5445-5455.2005

Chan, K.-L., Gligoris, T., Upcher, W., Kato, Y., Shirahige, K., Nasmyth, K., Beckouët, F., 2013. Pds5 promotes and protects cohesin acetylation. Proc. Natl. Acad. Sci. U.S.A. 110, 13020–13025. doi:10.1073/pnas.1306900110

Chan, K.-L., Roig, M.B., Hu, B., Beckouët, F., Metson, J., Nasmyth, K., 2012. Cohesin's DNA exit gate is distinct from its entrance gate and is regulated by acetylation. Cell 150, 961–974. doi:10.1016/j.cell.2012.07.028

Chao, W.C.H., Murayama, Y., Muñoz, S., Costa, A., Uhlmann, F., Singleton, M.R., 2015. Structural Studies Reveal the Functional Modularity of the Scc2-Scc4 Cohesin Loader. CellReports 12, 719–725. doi:10.1016/j.celrep.2015.06.071

Chatterjee, A., Zakian, S., Hu, X.-W., Singleton, M.R., 2013. Structural insights into the regulation of cohesion establishment by Wpl1. EMBO J. doi:10.1038/emboj.2013.16

Chiu, A., Revenkova, E., Jessberger, R., 2004. DNA Interaction and Dimerization of Eukaryotic SMC Hinge Domains. Journal of Biological Chemistry 279, 26233–26242. doi:10.1074/jbc.M402439200

Cimprich, K.A., Cortez, D., 2008. ATR: an essential regulator of genome integrity. Nat Rev Mol Cell Biol 9, 616–627. doi:10.1038/nrm2450

Ciosk, R., Shirayama, M., Shevchenko, A., Tanaka, T., Toth, A., Nasmyth, K., 2000. Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. Mol. Cell 5, 243–254.

Cobb, J.A., Bjergbaek, L., Shimada, K., Frei, C., Gasser, S.M., 2003. DNA polymerase stabilization at stalled replication forks requires Mec1 and the RecQ helicase Sgs1. EMBO J 22, 4325–4336. doi:10.1093/emboj/cdg391

Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., Gevaert, K., 2009. Improved visualization of protein consensus sequences by iceLogo. Nat. Methods 6, 786–787. doi:10.1038/nmeth1109-786

Cowtan, K., 2006. The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr. D Biol. Crystallogr. 62, 1002–1011. doi:10.1107/S0907444906022116

Crabbé, L., Thomas, A., Pantesco, V., De Vos, J., Pasero, P., Lengronne, A., 2010. Analysis of replication profiles reveals key role of RFC-Ctf18 in yeast replication stress response. Nat Struct Mol Biol 17, 1391–1397. doi:10.1038/nsmb.1932

Çamdere, G., Guacci, V., Stricklin, J., Koshland, D., 2015. The ATPases of cohesin interface with regulators to modulate cohesin-mediated DNA tethering. Elife 4. doi:10.7554/eLife.11315

Dorsett, D., Merkenschlager, M., 2013. Cohesin at active genes: a unifying theme for cohesin and gene expression from model organisms to humans. Curr. Opin.

Cell Biol. 1–7. doi:10.1016/j.ceb.2013.02.003
Elbatsh, A.M.O., Haarhuis, J.H.I., Petela, N., Chapard, C., Fish, A., Celie, P.H., Stadnik, M., Ristic, D., Wyman, C., Medema, R.H., Nasmyth, K., Rowland, B.D., 2016. Cohesin Releases DNA through Asymmetric ATPase- Driven Ring Opening. MOLCEL 61, 575–588. doi:10.1016/j.molcel.2016.01.025
Emsley, P., Cowtan, K., 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132. doi:10.1107/S0907444904019158
Evans, P., McCoy, A., 2007. An introduction to molecular replacement. Acta Crystallogr. D Biol. Crystallogr. 64, 1–10. doi:10.1107/S0907444907051554
Farina, A., Shin, J.H., Kim, D.H., Bermudez, V.P., Kelman, Z., Seo, Y.S., Hurwitz, J., 2008. Studies with the Human Cohesin Establishment Factor, ChlR1: ASSOCIATION OF ChlR1 WITH Ctf18-RFC AND Fen1. J. Biol. Chem. 283, 20925–20936. doi:10.1074/jbc.M802696200
Figiel, M., Chon, H., Cerritelli, S.M., Cybulska, M., Crouch, R.J., Nowotny, M., 2011. The Structural and Biochemical Characterization of Human RNase H2 Complex Reveals the Molecular Basis for Substrate Recognition and Aicardi-Goutieres Syndrome Defects. Journal of Biological Chemistry 286, 10540–10550. doi:10.1074/jbc.M110.181974
Gaiser, F., Tan, S., Richmond, T.J., 2000. Novel dimerization fold of RAP30/RAP74 in human TFIIF at 1.7 Å resolution. Journal of Molecular Biology 302, 1119–1127. doi:10.1006/jmbi.2000.4110
Gandhi, R., Gillespie, P.J., Hirano, T., 2006. Human Wapl is a cohesin-binding protein that promotes sister-chromatid resolution in mitotic prophase. Current Biology 16, 2406–2417. doi:10.1016/j.cub.2006.10.061
Gao, H., Moss, D.L., Parke, C., Tatum, D., Lustig, A.J., 2014. The Ctf18RFC Clamp Loader Is Essential for Telomere Stability in Telomerase-Negative and mre11 Mutant Alleles. PLoS ONE 9, e88633–13. doi:10.1371/journal.pone.0088633
García-Rodríguez, L.J., De Piccoli, G., Marchesi, V., Jones, R.C., Edmondson, R.D., Labib, K., 2015. A conserved Polϵ binding module in Ctf18-RFC is required for S-phase checkpoint activation downstream of Mec1. Nucleic Acids Research. doi:10.1093/nar/gkv799
Geiger, S.R., Lorenzen, K., Schreieck, A., Hanecker, P., Kostrewa, D., Heck, A.J.R., Cramer, P., 2010. RNA Polymerase I Contains a TFIIF-Related DNA-Binding Subcomplex. MOLCEL 39, 583–594. doi:10.1016/j.molcel.2010.07.028
Gerlich, D., Koch, B., Dupeux, F., Peters, J.-M., Ellenberg, J., 2006. Live-cell imaging reveals a stable cohesin-chromatin interaction after but not before DNA replication. Current Biology 16, 1571–1578. doi:10.1016/j.cub.2006.06.068
Gillis, L.A., McCallum, J., Kaur, M., DeScipio, C., Yaeger, D., Mariani, A., Kline, A.D., Li, H.-H., Devoto, M., Jackson, L.G., Krantz, I.D., 2004. NIPBL mutational analysis in 120 individuals with Cornelia de Lange syndrome and evaluation of genotype-phenotype correlations. Am. J. Hum. Genet. 75, 610–623. doi:10.1086/424698
Gligoris, T.G., Scheinost, J.C., Bürmann, F., Petela, N., Chan, K.-L., Uluocak, P., Beckouët, F., Gruber, S., Nasmyth, K., Löwe, J., 2014. Closing the cohesin ring: structure and function of its Smc3-kleisin interface. Science 346, 963–967. doi:10.1126/science.1256917
Gomes, X.V., Schmidt, S.L., Burgers, P.M., 2001. ATP utilization by yeast replication factor C. II. Multiple stepwise ATP binding events are required to load proliferating cell nuclear antigen onto primed DNA. J. Biol. Chem. 276, 34776–

34783. doi:10.1074/jbc.M011743200
Gordillo, M., Vega, H., Trainer, A.H., Hou, F., Sakai, N., Luque, R., Kayserili, H., Basaran, S., Skovby, F., Hennekam, R.C.M., Uzielli, M.L.G., Schnur, R.E., Manouvrier, S., Chang, S., Blair, E., Hurst, J.A., Forzano, F., Meins, M., Simola, K.O.J., Raas-Rothschild, A., Schultz, R.A., McDaniel, L.D., Ozono, K., Inui, K., Zou, H., Jabs, E.W., 2008. The molecular mechanism underlying Roberts syndrome involves loss of ESCO2 acetyltransferase activity. Human Molecular Genetics 17, 2172–2180. doi:10.1093/hmg/ddn116
Gorr, I.H., Boos, D., Stemmann, O., 2005. Mutual inhibition of separase and Cdk1 by two-step complex formation. MOLCEL 19, 135–141. doi:10.1016/j.molcel.2005.05.022
Green, C.M., Erdjument-Bromage, H., Tempst, P., Lowndes, N.F., 2000. A novel Rad24 checkpoint protein complex closely related to replication factor C. Current Biology 10, 39–42.
Guacci, V., Koshland, D., & Strunnikov, A. (1997). A direct link between sister chromatid cohesion and chromosome condensation revealed through the analysis of MCD1 in S. cerevisiae. Cell, 91(1), 47–57.
Gruber, S., Haering, C.H., Nasmyth, K., 2003. Chromosomal Cohesin Forms a Ring. Cell 112, 765–777. doi:10.1016/s0092-8674(03)00162-4
Gulbis, J.M., Kazmirski, S.L., Finkelstein, J., Kelman, Z., O'Donnell, M., Kuriyan, J., 2004. Crystal structure of the chi:psi subassembly of the Escherichia coli DNA polymerase clamp-loader complex. Eur J Biochem 271, 439–449. doi:10.1046/j.1432-1033.2003.03944.x
Haering, C.H., Löwe, J., Hochwagen, A., Nasmyth, K., 2002. Molecular architecture of SMC proteins and the yeast cohesin complex. MOLCEL 9, 773–788.
Haering, C.H., Schoffnegger, D., Nishino, T., Helmhart, W., Nasmyth, K., Löwe, J., 2004. Structure and stability of cohesin's Smc1-kleisin interaction. MOLCEL 15, 951–964. doi:10.1016/j.molcel.2004.08.030
Hahn, T., 2005. International Tables for Crystallography, Space-Group Symmetry.
Hanna, J.S., Kroll, E.S., Lundblad, V., Spencer, F.A., 2001. Saccharomyces cerevisiae CTF18 and CTF4 are required for sister chromatid cohesion. Mol. Cell. Biol. 21, 3144–3158. doi:10.1128/MCB.21.9.3144-3158.2001
Hara, K., Zheng, G., Qu, Q., Liu, H., Ouyang, Z., Chen, Z., Tomchick, D.R., Yu, H., 2014. Structure of cohesin subcomplex pinpoints direct shugoshin-Wapl antagonism in centromeric cohesion. Nat Struct Mol Biol 21, 864–870. doi:10.1038/nsmb.2880
Harami, G.M., Gyimesi, M., Kovács, M., 2013. From keys to bulldozers: expanding roles for winged helix domains in nucleic-acid-binding proteins. Trends in Biochemical Sciences 38, 364–371. doi:10.1016/j.tibs.2013.04.006
Hartman, T., Stead, K., Koshland, D., Guacci, V., 2000. Pds5p is an essential chromosomal protein required for both sister chromatid cohesion and condensation in Saccharomyces cerevisiae. The Journal of Cell Biology 151, 613–626.
Hauf, S., Roitinger, E., Koch, B., Dittrich, C.M., Mechtler, K., Peters, J.-M., 2005. Dissociation of cohesin from chromosome arms and loss of arm cohesion during early mitosis depends on phosphorylation of SA2. PLoS Biol. 3, e69. doi:10.1371/journal.pbio.0030069
Hauf, S., Waizenegger, I.C., Peters, J.M., 2001. Cohesin cleavage by separase required for anaphase and cytokinesis in human cells. Science 293, 1320–1323. doi:10.1126/science.1061376

Hedglin, M., Kumar, R., Benkovic, S.J., 2013. Replication clamps and clamp loaders. Cold Spring Harb Perspect Biol 5, a010165–a010165. doi:10.1101/cshperspect.a010165

Heidinger-Pauli, J.M., Onn, I., Koshland, D., 2010. Genetic evidence that the acetylation of the Smc3p subunit of cohesin modulates its ATP-bound state to promote cohesion establishment in Saccharomyces cerevisiae. Genetics 185, 1249–1256. doi:10.1534/genetics.110.116871

Heidinger-Pauli, J.M., Unal, E., Guacci, V., Koshland, D., 2008. The kleisin subunit of cohesin dictates damage-induced cohesion. Mol. Cell 31, 47–56. doi:10.1016/j.molcel.2008.06.005

Heidinger-Pauli, J.M., Unal, E., Koshland, D., 2009. Distinct targets of the Eco1 acetyltransferase modulate cohesion in S phase and in response to DNA damage. Mol. Cell 34, 311–321. doi:10.1016/j.molcel.2009.04.008

Hinshaw, S.M., Makrantoni, V., Kerr, A., Marston, A.L., Harrison, S.C., 2015. Structural evidence for Scc4-dependent localization of cohesin loading. Elife 4, e06057. doi:10.7554/eLife.06057

Hirano, T., 2006. At the heart of the chromosome: SMC proteins in action. Nat Rev Mol Cell Biol 7, 311–322. doi:10.1038/nrm1909

Hirota, Y., Lahti, J.M., 2000. Characterization of the enzymatic activity of hChlR1, a novel human DNA helicase. Nucleic Acids Research 28, 917–924.

Holm, L., Rosenström, P., 2010. Dali server: conservation mapping in 3D. Nucleic Acids Research 38, W545–9. doi:10.1093/nar/gkq366

Hornig, N.C.D., Knowles, P.P., McDonald, N.Q., Uhlmann, F., 2002. The dual mechanism of separase regulation by securin. Current Biology 12, 973–982.

Horsfield, J.A., Anagnostou, S.H., Hu, J.K.H., Cho, K.H.Y., Geisler, R., Lieschke, G., Crosier, K.E., Crosier, P.S., 2007. Cohesin-dependent regulation of Runx genes. Development 134, 2639–2649. doi:10.1242/dev.002485

Huis in 't Veld, P.J., Herzog, F., Ladurner, R., Davidson, I.F., Piric, S., Kreidl, E., Bhaskara, V., Aebersold, R., Peters, J.-M., 2014. Characterization of a DNA exit gate in the human cohesin ring. Science 346, 968–972. doi:10.1126/science.1256904

Ireland, M., Donnai, D., Burn, J., 1993. Brachmann-de Lange syndrome. Delineation of the clinical phenotype. Am. J. Med. Genet. 47, 959–964. doi:10.1002/ajmg.1320470705

Ivanov, D., Nasmyth, K., 2005. A Topological Interaction between Cohesin Rings and a Circular Minichromosome. Cell 122, 849–860. doi:10.1016/j.cell.2005.07.018

Ivanov, D., Schleiffer, A., Eisenhaber, F., Mechtler, K., Haering, C.H., Nasmyth, K., 2002. Eco1 is a novel acetyltransferase that can acetylate proteins involved in cohesion. Curr. Biol. 12, 323–328.

Judge, C., 1973. A sibship with the pseudothalidomide syndrome and an association with Rh incompatibility. Med. J. Aust. 2, 280–281.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., Taatjes, D.J., Dekker, J., Young, R.A., 2010. Mediator and cohesin connect gene expression and chromatin architecture. Nature 467, 430–435. doi:10.1038/nature09380

Kenna, M.A., Skibbens, R.V., 2003. Mechanical link between cohesion establishment and DNA replication: Ctf7p/Eco1p, a cohesion establishment factor, associates with three different replication factor C complexes. Mol. Cell. Biol. 23, 2999–3007.

Khurshid, S., Saridakis, E., Govada, L., Chayen, N.E., 2014. Porous nucleating

agents for protein crystallization. Nat Protoc 9, 1621–1633. doi:10.1038/nprot.2014.109

Kitajima, T.S., Kawashima, S.A., Watanabe, Y., 2004. The conserved kinetochore protein shugoshin protects centromeric cohesion during meiosis. Nature 427, 510–517. doi:10.1038/nature02312

Kitajima, T.S., Sakuno, T., Ishiguro, K.-I., Iemura, S.-I., Natsume, T., Kawashima, S.A., Watanabe, Y., 2006. Shugoshin collaborates with protein phosphatase 2A to protect cohesin. Nature 441, 46–52. doi:10.1038/nature04663

Kouprina, N., Kroll, E., Kirillov, A., Bannikov, V., Zakharyev, V., Larionov, V., 1994. CHL12, a gene essential for the fidelity of chromosome transmission in the yeast Saccharomyces cerevisiae. Genetics 138, 1067–1079.

Kouprina, N., Tsouladze, A., Koryabin, M., Hieter, P., Spencer, F., Larionov, V., 1993. Identification and genetic mapping of CHL genes controlling mitotic chromosome transmission in yeast. Yeast 9, 11–19. doi:10.1002/yea.320090103

Kouznetsova, E., Kanno, T., Karlberg, T., Thorsell, A.-G., Wisniewska, M., Kursula, P., Sjögren, C., Schüler, H., 2016. Sister Chromatid Cohesion Establishment Factor ESCO1 Operates by Substrate-Assisted Catalysis. Structure. doi:10.1016/j.str.2016.03.021

Kubota, T., Myung, K., Donaldson, A.D., 2014. Is PCNA unloading the central function of the Elg1/ATAD5 replication factor C-like complex? Cell Cycle 12, 2570–2579. doi:10.4161/cc.25626

Kubota, T., Nishimura, K., Kanemaki, M.T., Donaldson, A.D., 2013. The Elg1 replication factor C-like complex functions in PCNA unloading during DNA replication. Mol. Cell 50, 273–280. doi:10.1016/j.molcel.2013.02.012

Kueng, S., Hegemann, B., Peters, B.H., Lipp, J.J., Schleiffer, A., Mechtler, K., Peters, J.-M., 2006. Wapl Controls the Dynamic Association of Cohesin with Chromatin. Cell 127, 955–967. doi:10.1016/j.cell.2006.09.040

Kurze, A., Michie, K.A., Dixon, S.E., Mishra, A., Itoh, T., Khalid, S., Strmecki, L., Shirahige, K., Haering, C.H., Löwe, J., Nasmyth, K., 2010. A positively charged channel within the Smc1/Smc3 hinge required for sister chromatid cohesion. EMBO J 30, 364–378. doi:10.1038/emboj.2010.315

Ladurner, R., Bhaskara, V., Huis in 't Veld, P.J., Davidson, I.F., Kreidl, E., Petzold, G., Peters, J.-M., 2014. Cohesin's ATPase activity couples cohesin loading onto DNA with Smc3 acetylation. Curr. Biol. 24, 2228–2237. doi:10.1016/j.cub.2014.08.011

Lau, O.D., Kundu, T.K., Soccio, R.E., Ait-Si-Ali, S., Khalil, E.M., Vassilev, A., Wolffe, A.P., Nakatani, Y., Roeder, R.G., Cole, P.A., 2000. HATs off: selective synthetic inhibitors of the histone acetyltransferases p300 and PCAF. Mol. Cell 5, 589–595.

Lee, B.-G., Roig, M.B., Jansma, M., Petela, N., Metson, J., Nasmyth, K., Löwe, J., 2016. Crystal Structure of the Cohesin Gatekeeper Pds5 and in Complex with Kleisin Scc1. CellReports 1–9. doi:10.1016/j.celrep.2016.02.020

Lee, J., Kitajima, T.S., Tanno, Y., Yoshida, K., Morita, T., Miyano, T., Miyake, M., Watanabe, Y., 2008. Unified mode of centromeric protection by shugoshin in mammalian oocytes and somatic cells. Nat Cell Biol 10, 42–52. doi:10.1038/ncb1667

Lehmann, A.R., 2005. The role of SMC proteins in the responses to DNA damage. DNA Repair 4, 309–314. doi:10.1016/j.dnarep.2004.07.009

Lengronne, A., Katou, Y., Mori, S., Yokobayashi, S., Kelly, G.P., Itoh, T., Watanabe, Y., Shirahige, K., Uhlmann, F., 2004. Cohesin relocation from sites of

chromosomal loading to places of convergent transcription. Nature 430, 573–578. doi:10.1038/nature02742

Lengronne, A., McIntyre, J., Katou, Y., Kanoh, Y., Hopfner, K.-P., Shirahige, K., Uhlmann, F., 2006. Establishment of sister chromatid cohesion at the S. cerevisiae replication fork. MOLCEL 23, 787–799. doi:10.1016/j.molcel.2006.08.018

Liu, H., Rankin, S., Yu, H., 2012. Phosphorylation-enabled binding of SGO1--PP2A to cohesin protects sororin and centromeric cohesion during mitosis. Nat Cell Biol 15, 40–49. doi:10.1038/ncb2637

Liu, J., Krantz, I.D., 2008. Cohesin and Human Disease. Annu. Rev. Genom. Human Genet. 9, 303–320. doi:10.1146/annurev.genom.9.081307.164211

Lopez-Serra, L., Kelly, G., Patel, H., Stewart, A., Uhlmann, F., 2014. The Scc2-Scc4 complex acts in sister chromatid cohesion and transcriptional regulation by maintaining nucleosome-free regions. Nat. Genet. 46, 1147–1151. doi:10.1038/ng.3080

Losada, A., Hirano, M., Hirano, T., 2002. Cohesin release is required for sister chromatid resolution, but not for condensin-mediated compaction, at the onset of mitosis. Genes & Development 16, 3004–3016. doi:10.1101/gad.249202

Losada, A., Hirano, M., Hirano, T., 1998. Identification of Xenopus SMC protein complexes required for sister chromatid cohesion. Genes & Development 12, 1986–1997.

Losada, A., Yokochi, T., Kobayashi, R., Hirano, T., 2000. Identification and characterization of SA/Scc3p subunits in the Xenopus and human cohesin complexes. The Journal of Cell Biology 150, 405–416.

Lou, H., Komata, M., Katou, Y., Guan, Z., Reis, C.C., Budd, M., Shirahige, K., Campbell, J.L., 2008. Mrc1 and DNA polymerase epsilon function together in linking DNA replication and the S phase checkpoint. Mol. Cell 32, 106–117. doi:10.1016/j.molcel.2008.08.020

Lyons, N.A., Morgan, D.O., 2011. Cdk1-dependent destruction of Eco1 prevents cohesion establishment after S phase. Mol. Cell 42, 378–389. doi:10.1016/j.molcel.2011.03.023

Majka, J., Niedziela-Majka, A., Burgers, P.M.J., 2006. The checkpoint clamp activates Mec1 kinase during initiation of the DNA damage checkpoint. MOLCEL 24, 891–901. doi:10.1016/j.molcel.2006.11.027

Manuel García-Ruiz, J., 2003. Nucleation of protein crystals. Journal of Structural Biology 142, 22–31. doi:10.1016/S1047-8477(03)00035-2

Matthews, B.W., 1968. Solvent content of protein crystals. Journal of Molecular Biology 33, 491–497. doi:10.1016/0022-2836(68)90205-2

Mayer, M.L., Gygi, S.P., Aebersold, R., Hieter, P., 2001. Identification of RFC(Ctf18p, Ctf8p, Dcc1p): an alternative RFC complex required for sister chromatid cohesion in S. cerevisiae. Mol. Cell 7, 959–970. doi:10.1091/mbc.E03-08-0619

Mayer, M.L., Pot, I., Chang, M., Xu, H., Aneliunas, V., Kwok, T., Newitt, R., Aebersold, R., Boone, C., Brown, G.W., Hieter, P., 2004. Identification of protein complexes required for efficient sister chromatid cohesion. Mol. Biol. Cell 15, 1736–1745. doi:10.1091/mbc.E03-08-0619

Mc Intyre, J., Muller, E.G.D., Weitzer, S., Snydsman, B.E., Davis, T.N., Uhlmann, F., 2007. In vivo analysis of cohesin architecture using FRET in the budding yeast Saccharomyces cerevisiae. EMBO J 26, 3783–3793. doi:10.1038/sj.emboj.7601793

McCoy, A.J., 2004. Liking likelihood. Acta Crystallogr. D Biol. Crystallogr. 60, 2169–

2183. doi:10.1107/S0907444904016038
McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. Phaser crystallographic software. J Appl Crystallogr 40, 658–674. doi:10.1107/S0021889807021206
McPherson, A., Cudney, B., 2014. Optimization of crystallization conditions for biological macromolecules. Acta Cryst (2014). F70, 1445-1467 [doi:10.1107/S2053230X14019670] 70, 1–23. doi:10.1107/S2053230X14019670
Melby, T.E., Ciampaglio, C.N., Briscoe, G., Erickson, H.P., 1998. The symmetrical structure of structural maintenance of chromosomes (SMC) and MukB proteins: long, antiparallel coiled coils, folded at a flexible hinge. The Journal of Cell Biology 142, 1595–1604.
Michaelis, C., Ciosk, R., Nasmyth, K., 1997. Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. Cell 91, 35–45.
Moldovan, G.-L., Pfander, B., Jentsch, S., 2007. PCNA, the maestro of the replication fork. Cell 129, 665–679. doi:10.1016/j.cell.2007.05.003
Moldovan, G.-L., Pfander, B., Jentsch, S., 2006. PCNA controls establishment of sister chromatid cohesion during S phase. Mol. Cell 23, 723–732. doi:10.1016/j.molcel.2006.07.007
Morgan, D.O., 1997. Cyclin-dependent kinases: engines, clocks, and microprocessors. Annu. Rev. Cell Dev. Biol. 13, 261–291. doi:10.1146/annurev.cellbio.13.1.261
Muir, K.W., Kschonsak, M., Li, Y., Metz, J., Haering, C.H., Panne, D., 2016. Structure of the Pds5-Scc1 Complex and Implications for Cohesin Function. CellReports 0, 1–12. doi:10.1016/j.celrep.2016.01.078
Murakami, T., Takano, R., Takeo, S., Taniguchi, R., Ogawa, K., Ohashi, E., Tsurimoto, T., 2010. Stable interaction between the human proliferating cell nuclear antigen loader complex Ctf18-replication factor C (RFC) and DNA polymerase {epsilon} is mediated by the cohesion-specific subunits, Ctf18, Dcc1, and Ctf8. Journal of Biological Chemistry 285, 34608–34615. doi:10.1074/jbc.M110.166710
Murayama, Y., Uhlmann, F., 2015. DNA Entry into and Exit out of the Cohesin Ring by an Interlocking Gate Mechanism. Cell 163, 1628–1640. doi:10.1016/j.cell.2015.11.030
Murayama, Y., Uhlmann, F., 2013. Biochemical reconstitution of topological DNA binding by the cohesin ring. Nature 1–16. doi:10.1038/nature12867
Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., Vagin, A.A., 2011. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. D Biol. Crystallogr. 67, 355–367. doi:10.1107/S0907444911001314
Nasmyth, K., Haering, C.H., 2009. Cohesin: Its Roles and Mechanisms. Annu. Rev. Genet. 43, 525–558. doi:10.1146/annurev-genet-102108-134233
Navadgi-Patil, V.M., Burgers, P.M., 2008. Yeast DNA replication protein Dpb11 activates the Mec1/ATR checkpoint kinase. J. Biol. Chem. 283, 35853–35859. doi:10.1074/jbc.M807435200
Neuwald, A.F., Hirano, T., 2000. HEAT repeats associated with condensins, cohesins, and other complexes involved in chromosome-related functions. Genome Res. 10, 1445–1452. doi:10.1101/gr.147400
Nishiyama, T., Ladurner, R., Schmitz, J., Kreidl, E., Schleiffer, A., Bhaskara, V., Bando, M., Shirahige, K., Hyman, A.A., Mechtler, K., Peters, J.-M., 2010. Sororin

Mediates Sister Chromatid Cohesion by Antagonizing Wapl. Cell 143, 737–749. doi:10.1016/j.cell.2010.10.031

Ocampo-Hafalla, M.T., Katou, Y., Shirahige, K., Uhlmann, F., 2007. Displacement and re-accumulation of centromeric cohesin during transient pre-anaphase centromere splitting. Chromosoma 116, 531–544. doi:10.1007/s00412-007-0118-4

Okimoto, H., Tanaka, S., Araki, H., Ohashi, E., Tsurimoto, T., 2016. Conserved interaction of Ctf18-RFC with DNA polymerase ε is critical for maintenance of genome stability in saccharomyces cerevisiae. Genes Cells n/a–n/a. doi:10.1111/gtc.12356

Onn, I., Guacci, V., Koshland, D.E., 2009. The zinc finger of Eco1 enhances its acetyltransferase activity during sister chromatid cohesion. Nucleic Acids Research 37, 6126–6134. doi:10.1093/nar/gkp656

Ouyang, Z., Zheng, G., Song, J., Borek, D.M., Otwinowski, Z., Brautigam, C.A., Tomchick, D.R., Rankin, S., Yu, H., 2013. Structure of the human cohesin inhibitor Wapl. Proc. Natl. Acad. Sci. U.S.A. doi:10.1073/pnas.1304594110

Ouyang, Z., Zheng, G., Tomchick, D.R., Luo, X., Yu, H., 2016. Structural Basis and IP6 Requirement for Pds5-Dependent Cohesin Dynamics. MOLCEL 1–13. doi:10.1016/j.molcel.2016.02.033

Palecek, J.J., Gruber, S., 2015. Kite Proteins: a Superfamily of SMC/Kleisin Partners Conserved Across Bacteria, Archaea, and Eukaryotes. Structure/Folding and Design 23, 1–8. doi:10.1016/j.str.2015.10.004

Pan, X., Ye, P., Yuan, D.S., Wang, X., Bader, J.S., Boeke, J.D., 2006. A DNA integrity network in the yeast Saccharomyces cerevisiae. Cell 124, 1069–1081. doi:10.1016/j.cell.2005.12.036

Petronczki, M., 2004. Sister-chromatid cohesion mediated by the alternative RF-CCtf18/Dcc1/Ctf8, the helicase Chl1 and the polymerase- -associated protein Ctf4 is essential for chromatid disjunction during meiosis II. Journal of Cell Science 117, 3547–3559. doi:10.1242/jcs.01231

Pflugrath, J.W., 2015. IYCr crystallization series Practical macromolecular cryocrystallography. Acta Cryst (2015). F71, 622-642 [doi:10.1107/S2053230X15008304] 71, 1–21. doi:10.1107/S2053230X15008304

Poux, A.N., Cebrat, M., Kim, C.M., Cole, P.A., Marmorstein, R., 2002. Structure of the GCN5 histone acetyltransferase bound to a bisubstrate inhibitor. Proc. Natl. Acad. Sci. U.S.A. 99, 14065–14070. doi:10.1073/pnas.222373899

Putnam, C.D., Jaehnig, E.J., Kolodner, R.D., 2009. Perspectives on the DNA damage and replication checkpoint responses in Saccharomyces cerevisiae. DNA Repair 8, 974–982. doi:10.1016/j.dnarep.2009.04.021

Remeseiro, S., Cuadrado, A., Kawauchi, S., Calof, A.L., Lander, A.D., Losada, A., 2013a. Reduction of Nipbl impairs cohesin loading locally and affects transcription but not cohesion-dependent functions in a mouse model of Cornelia de Lange Syndrome. BBA - Molecular Basis of Disease 1832, 2097–2102. doi:10.1016/j.bbadis.2013.07.020

Remeseiro, S., Cuadrado, A., Losada, A., 2013b. Cohesin in development and disease. Development 140, 3715–3718. doi:10.1242/dev.090605

Rhodes, G., 2010. Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models.

Roig, M.B., Löwe, J., Chan, K.-L., Beckouët, F., Metson, J., Nasmyth, K., 2014. Structure and function of cohesin's Scc3/SA regulatory subunit. FEBS Letters

588, 3692–3702. doi:10.1016/j.febslet.2014.08.015
Rojowska, A., Lammens, K., Seifert, F.U., Direnberger, C., Feldmann, H., Hopfner, K.-P., 2014. Structure of the Rad50 DNA double-strand break repair protein in complex with DNA. EMBO J 33, 2847–2859. doi:10.15252/embj.201488889
Rowland, B.D., Roig, M.B., Nishino, T., Kurze, A., Uluocak, P., Mishra, A., Beckouët, F., Underwood, P., Metson, J., Imre, R., Mechtler, K., Katis, V.L., Nasmyth, K., 2009. Building Sister Chromatid Cohesion: Smc3 Acetylation Counteracts an Antiestablishment Activity. Mol. Cell 33, 763–774. doi:10.1016/j.molcel.2009.02.028
Saccardo, P., Corchero, J.L., Ferrer-Miralles, N., 2016. Tools to cope with difficult-to-express proteins. Appl Microbiol Biotechnol 100, 4347–4355. doi:10.1007/s00253-016-7514-8
Saitoh, N., Goldberg, I.G., Wood, E.R., Earnshaw, W.C., 1994. ScII: an abundant chromosome scaffold protein is a member of a family of putative ATPases with an unusual predicted tertiary structure. The Journal of Cell Biology 127, 303–318.
Salic, A., Waters, J.C., Mitchison, T.J., 2004. Vertebrate shugoshin links sister centromere cohesion and kinetochore microtubule stability in mitosis. Cell 118, 567–578. doi:10.1016/j.cell.2004.08.016
Samora, C.P., Saksouk, J., Goswami, P., Wade, B.O., Singleton, M.R., Bates, P.A., Lengronne, A., Costa, A., Uhlmann, F., 2016. Ctf4 Links DNA Replication with Sister Chromatid Cohesion Establishment by Recruiting the Chl1 Helicase to the Replisome. Mol. Cell. doi:10.1016/j.molcel.2016.05.036
Schüle, B., Oviedo, A., Johnston, K., Pai, S., Francke, U., 2005. Inactivating mutations in ESCO2 cause SC phocomelia and Roberts syndrome: no phenotype-genotype correlation. Am. J. Hum. Genet. 77, 1117–1128. doi:10.1086/498695
Seitan, V.C., Banks, P., Laval, S., Majid, N.A., Dorsett, D., Rana, A., Smith, J., Bateman, A., Krpic, S., Hostert, A., Rollins, R.A., Erdjument-Bromage, H., Tempst, P., Benard, C.Y., Hekimi, S., Newbury, S.F., Strachan, T., 2006. Metazoan Scc4 homologs link sister chromatid cohesion to cell and axon migration guidance. PLoS Biol. 4, e242. doi:10.1371/journal.pbio.0040242
Shaban, N.M., Harvey, S., Perrino, F.W., Hollis, T., 2010. The Structure of the Mammalian RNase H2 Complex Provides Insight into RNA{middle dot}DNA Hybrid Processing to Prevent Immune Dysfunction. Journal of Biological Chemistry 285, 3617–3624. doi:10.1074/jbc.M109.059048
Shimada, K., Pasero, P., Gasser, S.M., 2002. ORC and the intra-S-phase checkpoint: a threshold regulates Rad53p activation in S phase. Genes & Development 16, 3236–3252. doi:10.1101/gad.239802
Shintomi, K., Hirano, T., 2009. Releasing cohesin from chromosome arms in early mitosis: opposing actions of Wapl-Pds5 and Sgo1. Genes & Development 23, 2224–2236. doi:10.1101/gad.1844309
Shiomi, Y., Shinozaki, A., Sugimoto, K., Usukura, J., Obuse, C., Tsurimoto, T., 2004. The reconstituted human Chl12-RFC complex functions as a second PCNA loader. Genes Cells 9, 279–290. doi:10.1111/j.1356-9597.2004.00724.x
Simon, A.C., Zhou, J.C., Perera, R.L., van Deursen, F., Evrin, C., Ivanova, M.E., Kilkenny, M.L., Renault, L., Kjaer, S., Matak-Vinković, D., Labib, K., Costa, A., Pellegrini, L., 2014. A Ctf4 trimer couples the CMG helicase to DNA polymerase a in the eukaryotic replisome. Nature 510, 293–297. doi:10.1038/nature13234
Simonetta, K.R., Kazmirski, S.L., Goedken, E.R., Cantor, A.J., Kelch, B.A., McNally, R., Seyedin, S.N., Makino, D.L., O'Donnell, M., Kuriyan, J., 2009. The

mechanism of ATP-dependent primer-template recognition by a clamp loader complex. Cell 137, 659–671. doi:10.1016/j.cell.2009.03.044

Sjögren, C., Nasmyth, K., 2001. Sister chromatid cohesion is required for postreplicative double-strand break repair in Saccharomyces cerevisiae. Current Biology 11, 991–995.

Solomon, D.A., Kim, J.-S., Bondaruk, J., Shariat, S.F., Wang, Z.-F., Elkahloun, A.G., Ozawa, T., Gerard, J., Zhuang, D., Zhang, S., Navai, N., Siefker-Radtke, A., Phillips, J.J., Robinson, B.D., Rubin, M.A., Volkmer, B., Hautmann, R., Küfer, R., Hogendoorn, P.C.W., Netto, G., Theodorescu, D., James, C.D., Czerniak, B., Miettinen, M., Waldman, T., 2013. Frequent truncating mutations of STAG2 in bladder cancer. Nat. Genet. 45, 1428–1430. doi:10.1038/ng.2800

Spencer, F., Gerring, S. L., Connelly, C., & Hieter, P. (1990). Mitotic chromosome transmission fidelity mutants in Saccharomyces cerevisiae. *Genetics*, *124*(2), 237–249.

Sumara, I., Vorlaufer, E., Gieffers, C., Peters, B.H., Peters, J.M., 2000. Characterization of vertebrate cohesin complexes and their regulation in prophase. The Journal of Cell Biology 151, 749–762.

Sumara, I., Vorlaufer, E., Stukenberg, P.T., Kelm, O., Redemann, N., Nigg, E.A., Peters, J.-M., 2002. The dissociation of cohesin from chromosomes in prophase is regulated by Polo-like kinase. MOLCEL 9, 515–525.

Sun, P.D., Radaev, S., Kattah, M., IUCr, 2002. Generating isomorphous heavy-atom derivatives by a quick-soak method. Part I: test cases. Acta Crystallogr. D Biol. Crystallogr. 58, 1092–1098. doi:10.1107/S0907444902006510

Takahashi, T.S., Basu, A., Bermudez, V., Hurwitz, J., Walter, J.C., 2008. Cdc7-Drf1 kinase links chromosome cohesion to the initiation of DNA replication in Xenopus egg extracts. Genes & Development 22, 1894–1905. doi:10.1101/gad.1683308

Taylor, G.L., 2010. Introduction to phasing. Acta Cryst (2010). D66, 325-338 [doi:10.1107/S0907444910006694] 66, 1–14. doi:10.1107/S0907444910006694

Taylor, N.M.I., Baudin, F., Scheven, von, G., Muller, C.W., 2013. RNA polymerase III-specific general transcription factor IIIC contains a heterodimer resembling TFIIF Rap30/Rap74. Nucleic Acids Research 41, 9183–9196. doi:10.1093/nar/gkt664

Toth, A., Ciosk, R., Uhlmann, F., Galova, M., Schleiffer, A., Nasmyth, K., 1999. Yeast Cohesin complex requires a conserved protein, Eco1p(Ctf7), to establish cohesion between sister chromatids during DNA replication. Genes & Development 13, 320–333. doi:10.1101/gad.13.3.320

Uhlmann, F., 2009. A matter of choice: the establishment of sister chromatid cohesion. EMBO Rep 10, 1095–1102. doi:10.1038/embor.2009.207

Uhlmann, F., Wernic, D., Poupart, M.A., Koonin, E.V., Nasmyth, K., 2000. Cleavage of cohesin by the CD clan protease separin triggers anaphase in yeast. Cell 103, 375–386.

Unal, E., Heidinger-Pauli, J.M., Kim, W., Guacci, V., Onn, I., Gygi, S.P., Koshland, D.E., 2008. A Molecular Determinant for the Establishment of Sister Chromatid Cohesion. Science 321, 566–569. doi:10.1126/science.1157880

Unal, E., Heidinger-Pauli, J.M., Koshland, D., 2007. DNA Double-Strand Breaks Trigger Genome-Wide Sister-Chromatid Cohesion Through Eco1 (Ctf7). Science 317, 245–248. doi:10.1126/science.1140637

van den Ent, F., Löwe, J., 2006. RF cloning: A restriction-free method for inserting target genes into plasmids. Journal of Biochemical and Biophysical Methods 67,

67–74. doi:10.1016/j.jbbm.2005.12.008
Vaur, S., Feytout, A., Vazquez, S., Javerzat, J.-P., 2012. Pds5 promotes cohesin acetylation and stable cohesin-chromosome interaction. EMBO Rep 13, 645–652. doi:10.1038/embor.2012.72
Vega, H., Trainer, A.H., Gordillo, M., Crosier, M., Kayserili, H., Skovby, F., Uzielli, M.L.G., Schnur, R.E., Manouvrier, S., Blair, E., Hurst, J.A., Forzano, F., Meins, M., Simola, K.O.J., Raas-Rothschild, A., Hennekam, R.C.M., Jabs, E.W., 2010. Phenotypic variability in 49 cases of ESCO2 mutations, including novel missense and codon deletion in the acetyltransferase domain, correlates with ESCO2 expression and establishes the clinical criteria for Roberts syndrome. J. Med. Genet. 47, 30–37. doi:10.1136/jmg.2009.068395
Vernì, F., Gandhi, R., Goldberg, M.L., Gatti, M., 2000. Genetic and molecular analysis of wings apart-like (wapl), a gene controlling heterochromatin organization in Drosophila melanogaster. Genetics 154, 1693–1710.
Waizenegger, I.C., Hauf, S., Meinke, A., Peters, J.M., 2000. Two distinct pathways remove mammalian cohesin from chromosome arms in prophase and from centromeres in anaphase. Cell 103, 399–410.
Walden, H., 2010. Selenium incorporation using recombinant techniques. Acta Cryst (2010). D66, 352-357 [doi:10.1107/S0907444909038207] 66, 1–6. doi:10.1107/S0907444909038207
Walsh, M.A., Evans, G., Sanishvili, R., Dementieva, I., Joachimiak, A., 1999. MAD data collection - current trends. Acta Cryst (1999). D55, 1726-1732 [doi:10.1107/S0907444999008392] 55, 1–7. doi:10.1107/S0907444999008392
Warren, C.D., Eckley, D.M., Lee, M.S., Hanna, J.S., Hughes, A., Peyser, B., Jie, C., Irizarry, R., Spencer, F.A., 2004. S-phase checkpoint genes safeguard high-fidelity sister chromatid cohesion. Mol. Biol. Cell 15, 1724–1735. doi:10.1091/mbc.E03-09-0637
Xu, H., Boone, C., Brown, G.W., 2007. Genetic dissection of parallel sister-chromatid cohesion pathways. Genetics 176, 1417–1429. doi:10.1534/genetics.107.072876
Yamaguchi, K., Yamaguchi, R., Takahashi, N., Ikenoue, T., Fujii, T., Shinozaki, M., Tsurita, G., Hata, K., Niida, A., Imoto, S., Miyano, S., Nakamura, Y., Furukawa, Y., 2014. Overexpression of cohesion establishment factor DSCC1 through E2F in colorectal cancer. PLoS ONE 9, e85750. doi:10.1371/journal.pone.0085750
Zhang, J., Shi, X., Li, Y., Kim, B.-J., Jia, J., Huang, Z., Yang, T., Fu, X., Jung, S.Y., Wang, Y., Zhang, P., Kim, S.-T., Pan, X., Qin, J., 2008. Acetylation of Smc3 by Eco1 Is Required for S Phase Sister Chromatid Cohesion in Both Human and Yeast. Mol. Cell 31, 143–151. doi:10.1016/j.molcel.2008.06.006
Zhang, N., Panigrahi, A.K., Mao, Q., Pati, D., 2011. Interaction of Sororin protein with polo-like kinase 1 mediates resolution of chromosomal arm cohesion. Journal of Biological Chemistry 286, 41826–41837. doi:10.1074/jbc.M111.305888